Supplementary Information

Constrained Multistate Sequence Design for Nucleic Acid Reaction Pathway Engineering

Brian R. Wolfe[1,#], Nicholas J. Porubsky[2,#], Joseph N. Zadeh[1], Robert M. Dirks[1,‡], and Niles A. Pierce[1,3,4,*]

Contents

[1]Division of Biology & Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [2]Division of Chemistry & Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [3]Division of Engineering & Applied Science, California Institute of Technology, Pasadena, CA 91125, USA. [4]Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, UK. [#]These authors contributed equally. [‡]Deceased. [*]Corresponding author: niles@caltech.edu

List of Figures

List of Tables

List of Algorithms

S1 Algorithm

The constrained multistate test tube design algorithm described in the present work builds on the test tube design algorithm described by Wolfe and Pierce.[1] The two algorithms were developed concurrently so that the notation and concepts employed for sequence design over the ensemble of a single test tube would generalize naturally to performing sequence design over the ensemble of an arbitrary number of test tubes subject to user-specified sequence constraints. Readers interested in a detailed understanding of the present algorithm will benefit from reading the Algorithm section of Reference 1, which contains thorough descriptions of a subset of the algorithmic ingredients used in the present work. For conciseness, if an algorithmic ingredient requires no or minimal generalization and there is little risk of confusion, we simply refer to Reference 1 for details (using the same section heading for clarity). If some generalization in notation or concept is required and there is a risk of confusion, we restate the description of Reference 1 with updated details below (again using the same section heading for clarity). If a new algorithmic ingredient is required in the present setting, we provide full details below.

S1.1 Secondary Structure Model

The secondary structure, s, of one or more interacting nucleic acid strands is defined by a set of base pairs.[2] A *polymer graph* representation of a secondary structure is constructed by ordering the strands around a circle, drawing the backbones in succession from 5′ to 3′ around the circumference with a *nick* between each strand, and drawing straight lines connecting paired bases. A secondary structure is *unpseudoknotted* if there exists a strand ordering for which the polymer graph has no crossing lines. A secondary structure is *connected* if no subset of the strands is free of the others. A *complex* of interacting strands is specified as a strand ordering, π, corresponding to the *structural ensemble*, Γ, containing all connected polymer graphs with no crossing lines.[1] (We dispense with our prior convention[2–4] of calling this entity an *ordered complex*.) See Section S1.3 of Reference 1 for a discussion of distinguishability issues. A *test tube* may contain an arbitrary number of strand species interacting to form an arbitrary number of complex species in a dilute solution.

The sequence, ϕ, of a complex is specified as a list of bases $\phi^a \in \{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{U}\}$ for $a = 1, \ldots, |\phi|$ (T replaces U for DNA). Each base pair in a secondary structure is a Watson–Crick pair (A·U or C·G) or a wobble pair (G·U). For sequence ϕ and secondary structure $s \in \Gamma$, the *free energy*, $\Delta G(\phi, s)$, is calculated using nearest-neighbor empirical parameters for RNA[5–7] in 1M Na^+ or for DNA[7,8] in user-specified concentrations of Na^+ and Mg^{++}.[9,10] These physical models have practical utility for the analysis[11–19] and design[11,12,20–49] of functional nucleic acid systems, and provide the basis for rational analysis and design of equilibrium base-pairing in test tube ensembles for reaction pathway engineering.

S1.2 Analyzing Equilibrium Base-Pairing in the Multistate Test Tube Ensemble

Let Ψ_h^0 denote the set of strand species that interact in test tube $h \in \Omega$ to form the set of complex species Ψ_h. For complex $j \in \Psi_h$, with sequence ϕ_j and structural ensemble Γ_j, the partition function

$$Q(\phi_j) = \sum_{s \in \Gamma_j} \exp\left[-\Delta G(\phi_j, s)/k_B T\right]$$

can be used to calculate the equilibrium probability of any secondary structure $s \in \Gamma_j$:

$$p(\phi_j, s) = \exp\left[-\Delta G(\phi_j, s)/k_B T\right]/Q(\phi_j).$$

Here, k_B is the Boltzmann constant and T is temperature. The equilibrium base-pairing properties of complex j are characterized by the base-pairing probability matrix $P(\phi_j)$, with entries $P^{a,b}(\phi_j) \in [0, 1]$ corresponding to the probability,

$$P^{a,b}(\phi_j) = \sum_{s \in \Gamma_j} p(\phi_j, s) S^{a,b}(s),$$

that base pair $a \cdot b$ forms at equilibrium within ensemble Γ_j. Here, $S(s)$ is a *structure matrix* with entries $S^{a,b}(s) = 1$ if structure s contains base pair $a \cdot b$ and $S^{a,b}(s) = 0$ otherwise. For convenience, the structure and probability matrices are augmented with an extra column to describe unpaired bases. The entry $S^{a,|s|+1}(s)$ is unity if base a is unpaired in structure s and zero otherwise; the entry $P^{a,|\phi_j|+1}(\phi_j) \in [0,1]$ denotes the equilibrium probability that base a is unpaired over ensemble Γ_j. Hence the row sums of the augmented $S(s)$ and $P(\phi_j)$ matrices are unity.

Let $Q_{\Psi_h} \equiv Q_j \ \forall j \in \Psi_h$ denote the set of partition functions for the complexes in tube h. The set of equilibrium concentrations, x_{h,Ψ_h}, (specified as mole fractions) are the unique solution to the strictly convex optimization problem:[2]

$$\min_{x_{h,\Psi_h}} \sum_{j \in \Psi_h} x_{h,j}(\log x_{h,j} - \log Q_j - 1) \tag{S1a}$$

$$\text{subject to} \quad A_{i,j} x_{h,j} = x^0_{h,i} \quad \forall i \in \Psi^0_h, \tag{S1b}$$

where the constraints impose conservation of mass. A is the stoichiometry matrix with entries $A_{i,j}$ corresponding to the number of strands of type i in complex j, and $x^0_{h,i}$ is the total concentration of strand i introduced to test tube h.

To analyze the equilibrium base-pairing properties of all test tubes $h \in \Omega$, the partition function, Q_j, and equilibrium pair probability matrix, P_j, must be calculated for each complex $j \in \Psi$ using $\Theta(|\phi_j|^3)$ dynamic programs.[2,50–57] The equilibrium concentrations, x_{h,Ψ_h} $\forall h \in \Omega$, are calculated by solving a convex programming problem using an efficient trust region method at a cost that is typically negligible by comparison.[2] The overall time complexity to analyze the test tubes in Ω is then $O(|\Psi||\phi|^3_{\max})$, where $|\phi|_{\max}$ is the size of the largest complex.

Evaluation of the multistate test tube ensemble defect, $\mathcal{M}$, requires calculation of the complex partition functions, Q_Ψ, which are used to calculate the equilibrium concentrations, x_{h,Ψ_h} $\forall h \in \Omega$, as well as the equilibrium pair probability matrices, $P_{\Psi^{\text{on}}}$, which are used to calculate the complex ensemble defects, $n_{\Psi^{\text{on}}}$, and the normalized test tube ensemble defects, $\mathcal{M}_\Omega$. Hence, the time complexity to evaluate the design objective function over the set of target test tubes, Ω, is the same as the time complexity to analyze equilibrium base-pairing in Ω.

S1.3 Test Tube Ensemble Focusing

To reduce the cost of sequence optimization, the set of complexes, Ψ, is partitioned into two disjoint sets:

$$\Psi = \Psi^{\text{active}} \cup \Psi^{\text{passive}},$$

where Ψ^{active} denotes complexes that will be actively designed and Ψ^{passive} denotes complexes that will inherit sequence information from Ψ^{active}. Only the complexes in Ψ^{active} are directly accounted for in the focused test tube ensembles that are used to evaluate candidate sequences. Initially, we set

$$\Psi^{\text{active}} = \Psi^{\text{on}}, \quad \Psi^{\text{passive}} = \Psi^{\text{off}}, \tag{S2}$$

where

$$\Psi^{\text{on}} \equiv \cup_{h \in \Omega} \Psi^{\text{on}}_h$$

is the set of complexes that appear as on-targets in at least one test tube, and

$$\Psi^{\text{off}} \equiv \Psi - \Psi^{\text{on}}$$

is the set of complexes that appear as off-targets in at least one test tube and do not appear as on-targets in any test tube. Hence, with test tube ensemble focusing, only complexes that are on-targets in at least one test tube are actively designed at the outset of sequence design.

S1.4 Hierarchical Ensemble Decomposition

To enable efficient estimation of test tube ensemble properties, the structural ensemble Γ_j of each complex $j \in \Psi^{\text{active}}$ is hierarchically decomposed into a (possibly unbalanced) binary tree of conditional subensembles, yielding a forest of decomposition trees. Each complex $j \in \Psi^{\text{active}}$ contributes a single tree to the decomposition forest whether it is contained in one or more tubes $h \in \Omega$. The structural ensemble of each parent node within the forest is decomposed using one or more exclusive split-points to partition the parent nucleotides to its children. See Reference 1 for details on hierarchical ensemble decomposition. Let Λ denote the set of all nodes in the forest. Let Λ_d denote the set of all nodes at depth d.

S1.4.1 Structure-Guided Decomposition of On-Target Complexes

At the outset of sequence design, equilibrium base-pairing probabilities are not yet available to guide ensemble decomposition. Instead, structure-guided hierarchical ensemble decomposition is performed (using a single split-point per parent) for each on-target complex $j \in \Psi^{\text{active}}$, yielding a forest of $|\Psi^{\text{on}}|$ decomposition trees. See Reference 1 for details on structure-guided decomposition.

S1.4.2 Stop Condition Stringency

In order to build in a tolerance for a basal level of decomposition defect as subsequences are merged moving up the decomposition forest, the stringency of the stop condition (3) is increased by a factor of $f_{\text{stringent}} \in (0, 1)$ at each level moving down the decomposition forest:

$$f_d^{\text{stop}} \equiv f_{\text{stop}}(f_{\text{stringent}})^{d-1} \quad \forall d \in \{1, \ldots, D\}.$$

S1.5 Efficient Estimation of Test Tube Ensemble Properties

During sequence optimization, the design objective function is estimated based on physical quantities calculated efficiently at any depth $d \in \{1, \ldots, D\}$ in the decomposition forest.

S1.5.1 Complex Partition Function Estimate

For each complex $j \in \Psi^{\text{active}}$, the complex partition function estimate, $\tilde{Q}_j$, is calculated from conditional partition functions evaluated efficiently at any depth $d \in \{1, \ldots, D\}$ as described in Reference 1.

S1.5.2 Complex Pair Probability Matrix Estimate

For each complex $j \in \Psi^{\text{active}}$, the complex pair probability matrix estimate, $\tilde{P}_j$, is calculated from conditional pair probability matrices evaluated efficiently at any depth $d \in \{1, \ldots, D\}$ as described in Reference 1.

S1.5.3 Complex Concentration Estimate using Deflated Mass Constraints

For each tube $h \in \Omega$, the complex concentration estimates, $\tilde{x}_{h,\Psi_h^{\text{active}}}$, are calculated using the complex partition function estimates $\tilde{Q}_{\Psi_h^{\text{active}}}$ previously evaluated at any depth $d \in \{1, \ldots, D\}$ as described in Reference 1. Deflated mass constraints are used to model the effect of the neglected off-target complexes in Ψ_h^{passive}.

S1.5.4 Complex Ensemble Defect Estimate

For each complex $j \in \Psi^{\text{on}}$, the complex ensemble defect estimate, $\tilde{n}_j$, is calculated using the complex pair probability matrix estimate $\tilde{P}_j$ previously evaluated at any depth $d \in \{1, \ldots, D\}$ as described in Reference 1. For complex

j, the contribution of nucleotide a to the complex ensemble defect estimate is given by:

$$\tilde{n}_j^a = 1 - \sum_{1 \le b \le |\phi_j|+1} \tilde{P}_j^{a,b} S_j^{a,b}$$

and the complex ensemble defect estimate is then:

$$\tilde{n}_j = \sum_{1 \le a \le |\phi_j|} \tilde{n}_j^a. \tag{S3}$$

S1.5.5 Test Tube Ensemble Defect Estimate

For each tube $h \in \Omega$, the test tube ensemble defect estimate based on $\tilde{x}_{h,\Psi_h^{\text{active}}}$ and $\tilde{n}_{\Psi_h^{\text{on}}}$ calculated at any depth $d \in \{1, \ldots, D\}$, is:

$$\tilde{C}_h = \sum_{j \in \Psi_h^{\text{on}}} \tilde{c}_{h,j}, \tag{S4}$$

where

$$\tilde{c}_{h,j} = \tilde{n}_j \min\left(\tilde{x}_{h,j}, y_{h,j}\right) + |\phi_j| \max\left(y_{h,j} - \tilde{x}_{h,j}, 0\right) \tag{S5}$$

is the contribution of complex j. The normalized test tube ensemble defect estimate for tube $h \in \Omega$ at depth $d \in \{1, \ldots, D\}$ is then:

$$\tilde{\mathcal{M}}_h = \tilde{C}_h / y_h^{\text{nt}}, \tag{S6}$$

where

$$y_h^{\text{nt}} = \sum_{j \in \Psi_h^{\text{on}}} |\phi_j| y_{h,j}$$

is the total concentration of nucleotides in tube h.

S1.5.6 Multistate Test Tube Ensemble Defect Estimate

For the set of target test tubes Ω, the objective function estimate based on $\tilde{\mathcal{M}}_\Omega$ evaluated at any depth $d \in \{1, \ldots, D\}$ is then:

$$\tilde{\mathcal{M}} = \frac{1}{|\Omega|} \sum_{h \in \Omega} \tilde{\mathcal{M}}_h. \tag{S7}$$

We write $\tilde{\mathcal{M}}_d$ in subsequent equations where it is helpful note the depth d at which $\tilde{\mathcal{M}}$ was calculated.

Note that equations (S3)–(S7) may be collected into the single equation:

$$\tilde{\mathcal{M}} = \sum_{h \in \Omega} \sum_{j \in \Psi_h^{\text{on}}} \sum_{1 \le a \le |\phi_j|} \tilde{\mathcal{M}}_{h,j}^a \tag{S8}$$

where

$$\tilde{\mathcal{M}}_{h,j}^a \equiv \frac{1}{|\Omega| y_h^{\text{nt}}} \left[\tilde{n}_j^a \min(\tilde{x}_{h,j}, y_{h,j}) + \max(y_{h,j} - \tilde{x}_{h,j}, 0)\right], \tag{S9}$$

is the contribution of nucleotide a in complex $j \in \Psi_h^{\text{on}}$ in tube $h \in \Omega$ to the multistate test tube ensemble defect estimate, $\tilde{\mathcal{M}}$, evaluated at any depth $d \in \{1, \ldots, D\}$. This representation is convenient when defining objective function weights (Section S1.6) and when defining defect-weighted mutation sampling during leaf mutation and defect-weighted reseeding during leaf reoptimization (Sections S1.7.2 and S1.7.3).

S1.6 Adjusting Design Priorities using Defect Weights

The user may adjust design priorities by specifying weights for contributions to the multistate test tube ensemble defect estimate, $\tilde{\mathcal{M}}$:

- **Nucleotide weight:** $w_{h,j}^a$ weights the contribution of nucleotide a in complex $j \in \Psi_h^{\text{on}}$ in tube $h \in \Omega$.
- **Complex weight:** $w_{h,j}$ weights the contribution of complex $j \in \Psi_h^{\text{on}}$ in tube $h \in \Omega$ (equivalent to setting $w_{h,j}^a$ for all nucleotides $1 \leq a \leq |\phi_j|$).
- **Test tube weight:** w_h weights the contribution of tube $h \in \Omega$ (equivalent to setting $w_{h,j}$ for all complexes $j \in \Psi_h^{\text{on}}$).

Each weight takes a value in the interval $[0, \infty)$. By default, all weights are unity. Increasing the weight for a nucleotide, complex, or test tube will lead to a corresponding increase in the allocation of effort to designing this entity, typically leading to a corresponding reduction in the defect contribution of the entity. Likewise, decreasing the weight for a nucleotide, complex, or test tube will lead to a corresponding decrease in the allocation of effort to designing this entity, typically leading to a corresponding increase in the defect contribution of the entity.

Weights are incorporated into the objective function by replacing the defect contribution (S9) with the weighted defect contribution

$$\tilde{\mathcal{M}}_{h,j}^a \equiv \frac{w_h w_{h,j} w_{h,j}^a}{|\Omega| y_h^{\text{nt}}} \left[\tilde{n}_j^a \min(\tilde{x}_{h,j}, y_{h,j}) + \max(y_{h,j} - \tilde{x}_{h,j}, 0)\right], \tag{S10}$$

and summing using (S8) as before. If desired, the user can set weights at all three levels, leading to a multiplicative effect. The complex weights and test tube weights exist purely for convenience, as their effects can always be replicated by appropriately setting nucleotide weights (more tediously).

S1.7 Sequence Optimization at the Leaves of the Decomposition Forest

S1.7.1 Initialization

At the outset of sequence optimization, sequences are randomly initialized subject to the constraints in $\mathcal{R}$ by solving a constraint satisfaction problem using a branch and propagate algorithm (Section S1.11.3).

S1.7.2 Leaf Mutation

To minimize computational cost, all candidate mutation sets are evaluated at the leaf nodes, $k \in \Lambda_D$, of the decomposition forest. Leaf mutation terminates if the *leaf stop condition*,

$$\tilde{\mathcal{M}}_D \leq f_D^{\text{stop}}, \tag{S11}$$

is satisfied. Here, $\tilde{\mathcal{M}}_D$ denotes the objective function estimated at level D. A candidate mutation set is accepted if it decreases the objective function estimate (S8) and rejected otherwise.

We perform *defect weighted mutation sampling* by selecting nucleotide a in complex $j \in \Psi_h^{\text{on}}$ in tube $h \in \Omega$ for mutation with probability,

$$\tilde{\mathcal{M}}_{h,j}^a / \tilde{\mathcal{M}}_D, \tag{S12}$$

proportional to its contribution to the objective function. After selecting a candidate mutation position, a candidate mutation is randomly selected from the set of permitted nucleotides at that position. If the resulting sequence is infeasible (due to constraint violations caused by the candidate mutation), a feasible candidate sequence, $\hat{\phi}_{\Lambda_D}$, is generated by solving a constraint satisfaction problem using a branch and propagate algorithm (Section S1.11.4).

A feasible candidate sequence, $\hat{\phi}_{\Lambda_D}$, is evaluated via calculation of the objective function estimate, $\tilde{\mathcal{M}}_D$, if the candidate mutation set, ξ, is not in the set of previously rejected mutation sets, γ_{bad}. The set, γ_{bad}, is updated after each unsuccessful mutation and cleared after each successful mutation. The counter m_{bad} is used to keep track of the number of consecutive failed mutation attempts; it is incremented after each unsuccessful mutation and reset to zero after each successful mutation. Leaf mutation terminates unsuccessfully if $m_{\text{bad}} \geq M_{\text{bad}}$. The outcome of leaf mutation is the feasible sequence, ϕ_{Λ_D}, corresponding to the lowest encountered $\tilde{\mathcal{M}}_D$.

S1.7.3 Leaf Reoptimization

After leaf mutation terminates, if the leaf stop condition (S11) is not satisfied, leaf reoptimization commences. At the outset of each round of leaf reoptimization, we perform *defect-weighted reseeding* of M_{reseed} positions by selecting nucleotide a for reseeding (with a new random initial sequence) with probability (S12). Following reseeding, a feasible candidate sequence, $\hat{\phi}_{\Lambda_D}$, is generated by solving a constraint satisfaction problem using a branch and propagate algorithm (Section S1.11.5). After a new round of leaf mutation starting from this reseeded feasible sequence, the reoptimized candidate sequence, $\hat{\phi}_{\Lambda_D}$, is accepted if it decreases $\tilde{\mathcal{M}}_D$ and rejected otherwise. The counter m_{reopt} is used to keep track of the number of rounds of leaf reoptimization; m_{reopt} is incremented after each rejection and reset to zero after each acceptance. Leaf reoptimization terminates successfully if the leaf stop condition is satisfied and unsuccessfully if $m_{\text{reopt}} \geq M_{\text{reopt}}$. The outcome of leaf reoptimization is the feasible sequence, ϕ_{Λ_D}, corresponding to the lowest encountered $\tilde{\mathcal{M}}_D$.

S1.8 Subsequence Merging, Redecomposition, and Reoptimization

Moving down the decomposition forest, hierarchical ensemble decomposition makes the assumption that base pairs sandwiching parental split-points form with probability approaching unity. Conditional child ensembles enforce these sandwiching base pairs at all levels in the decomposition forest in accordance with the decomposition assumption. As subsequences are merged moving up the decomposition forest, the accuracy of the decomposition assumption is checked. If the assumption is correct, the child-estimated defect will accurately predict the parent-estimated defect. If the assumption is incorrect, the child-estimated defect will not accurately predict the parent-estimated defect since the conditional child ensembles neglect the contributions of structures that lack the sandwiching base pairs. During subsequence merging, if the decomposition assumption is discovered to be incorrect, hierarchical ensemble redecomposition is performed based on the newly available parental base-pairing information. The details of subsequence merging, redecomposition, and reoptimization are as follows.

After leaf reoptimization terminates, parent nodes at depth $d = D - 1$ merge their left and right child sequences to create the candidate sequence $\hat{\phi}_{\Lambda_d}$. The parental objective function estimate, $\tilde{\mathcal{M}}_d$, is calculated and the candidate sequence, $\hat{\phi}_{\Lambda_d}$, is accepted if it decreases $\tilde{\mathcal{M}}_d$ and rejected otherwise. If the *parental stop condition*

$$\tilde{\mathcal{M}}_d \leq \max(f_d^{\text{stop}}, \tilde{\mathcal{M}}_{d+1}/f_{\text{stringent}}) \tag{S13}$$

is satisfied, merging continues up to the next level in the forest. Otherwise, failure to satisfy the parental stop condition indicates the existence of the *decomposition defect*,

$$\tilde{\mathcal{M}}_d - \tilde{\mathcal{M}}_{d+1}/f_{\text{stringent}} > 0,$$

exceeding the basal level permitted by the parameter $f_{\text{stringent}}$. The parent node at depth d whose replacement by its children results in the greatest underestimate of the objective function at level d is subjected to structure- and probability-guided hierarchical ensemble decomposition (Section S1.10.2). Additional parents are redecomposed until

$$\tilde{\mathcal{M}}_d - \tilde{\mathcal{M}}^*_{d+1}/f_{\text{stringent}} \leq f_{\text{redecomp}}(\tilde{\mathcal{M}}_d - \tilde{\mathcal{M}}_{d+1}/f_{\text{stringent}})$$

where $\tilde{\mathcal{M}}_{d+1}$ is the child defect estimate before any redecomposition, $\tilde{\mathcal{M}}^*_{d+1}$ is the child defect estimate after redecomposition, and $f_{\text{redecomp}} \in (0, 1)$.

After redecomposition, the current sequences at depth d are pushed to level D, the lowest encountered defect estimate is reset for all levels below d, and a new round of leaf mutation and leaf reoptimization is performed. Following leaf reoptimization, merging begins again. Subsequence merging and reoptimization terminate successfully if the parental stop condition (S13) is satisfied at depth $d = 1$. The outcome of subsequence merging, redecomposition, and reoptimization is the feasible sequence, ϕ_{Λ_1}, corresponding to the lowest encountered $\tilde{\mathcal{M}}_1$.

S1.9 Test Tube Evaluation, Refocusing, and Reoptimization

Using test tube ensemble focusing, initial sequence optimization is performed for the on-target complexes in Ψ^{active}, neglecting the off-target complexes in Ψ^{passive}. At the termination of initial forest optimization, the estimated design objective function is $\tilde{\mathcal{M}}_1$, calculated using (S8). The estimated contributions for each tube $h \in \Omega$ are based on complex concentration estimates, $\tilde{x}_{h,\Psi_h^{\text{active}}}$, calculated using deflated total strand concentrations (equation (10) of Reference 1) to create a built-in defect allowance for the effect of the neglected off-targets in Ψ_h^{passive}. The exact design objective function, $\mathcal{M}$, is then evaluated for the first time over the full ensemble Ψ. For this exact calculation, the objective function, $\mathcal{M}$, is based on complex concentrations, x_{h,Ψ_h}, calculated using the full strand concentrations (equation (9) of Reference 1).

If the objective function satisfies the *termination stop condition*,

$$\mathcal{M} \leq \max(f_{\text{stop}}, \tilde{\mathcal{M}}_1), \tag{S14}$$

sequence design terminates successfully. Otherwise, failure to satisfy the termination stop condition indicates the existence of the *focusing defect*,

$$\mathcal{M} - \tilde{\mathcal{M}}_1 > 0. \tag{S15}$$

The multistate test tube ensemble is refocused by transferring the highest-concentration off-target in Ψ^{passive} to Ψ^{active}. Additional off-targets are transferred from Ψ^{passive} to Ψ^{active} until

$$\mathcal{M} - \tilde{\mathcal{M}}_1^* \leq f_{\text{refocus}}(\mathcal{M} - \tilde{\mathcal{M}}_1), \tag{S16}$$

where $\tilde{\mathcal{M}}_1$ is the forest-estimated defect before any refocusing, $\tilde{\mathcal{M}}_1^*$ is the forest-estimated defect after refocusing (calculated using deflated total strand concentrations (equation (10) of Reference 1) if $\Psi^{\text{passive}} \neq \emptyset$), and $f_{\text{refocus}} \in (0, 1)$.

The new off-targets in Ψ^{active} are then decomposed using probability-guided hierarchical ensemble decomposition (Section S1.10.1), the decomposition forest is augmented with new nodes at all depths, and forest reoptimization commences starting from the final sequences from the previous round of forest optimization. During forest reoptimization, the algorithm actively attempts to destabilize the off-targets that were added to Ψ^{active}. This process of test tube ensemble refocusing and forest reoptimization is repeated until the termination stop condition (S14) is satisfied, which is guaranteed to occur in the event that all off-targets are eventually added to Ψ^{active}. At the conclusion of sequence design, the algorithm returns the feasible sequence set, ϕ_Ψ, that yielded the lowest encountered objective function, $\mathcal{M}$.

S1.10 Hierarchical Ensemble Decomposition Using Multiple Exclusive Split-Points

Prior to sequence optimization, in the absence of base-pairing probability information, hierarchical ensemble decomposition is performed for each complex $j \in \Psi^{\text{active}}$ based on user-specified target structures. During subsequence merging, if decomposition defects are encountered, or during test tube evaluation, if focusing defects are encountered, subsequent hierarchical ensemble decomposition takes advantage of the newly available parental base-pairing probabilities. In either case, selection of the optimal set of exclusive split-points is determined using a branch and bound algorithm to minimize the cost of evaluating the child nodes (see Section S1.4 of Reference 1).

S1.10.1 Probability-Guided Decomposition using Multiple Exclusive Split-Points

During redecomposition (Section S1.8) and refocusing (Section S1.9), parent nodes that lack a target structure are decomposed via probability-guided decomposition using multiple exclusive split-points. See Reference 1 for details on probability-guided decomposition.

S1.10.2 Structure- and Probability-Guided Decomposition using Multiple Exclusive Split-Points

During redecomposition (Section S1.8), parent nodes that have a target structure are decomposed via structure- and probability-guided decomposition using multiple exclusive split-points. See Reference 1 for details on structure- and probability-guided decomposition.

Table S1. IUPAC degenerate nucleotide codes for RNA.

Code	Nucleotides
M	A or C
R	A or G
W	A or U
S	C or G
Y	C or U
K	G or U
V	A, C, or G
H	A, C, or U
D	A, G, or U
B	C, G, or U
N	A, C, G, or U

T replaces U for DNA.

S1.10.3 Multistate Test Tube Ensemble Defect Estimate Using Multiple Exclusive Decompositions

Because exclusive split-points lead to exclusive structural ensembles, the expressions used to estimate ensemble properties over Ω (Section S1.5) can be generalized to account for the possibility of multiple exclusive split-points within any parent in the decomposition forest. See Reference 1 for details.

S1.11 Generation of Feasible Sequences

Each time the sequence is initialized, mutated, or reseeded, a feasible sequence is generated by solving a constraint satisfaction problem based on the user-specified constraints in $\mathcal{R}$.

S1.11.1 Constraint Satisfaction Problem

A constraint satisfaction problem (CSP)[58] is specified as:

- a set of *variables*,
- a set of *domains*, each listing the possible values for the corresponding variable,
- a set of *constraints*, each defined by a constraint relation operating on a subset of the variables.

In the present setting, each variable is the sequence, ϕ^a, of a nucleotide, a. For RNA, the domain for each variable is $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{U}\}$. Each constraint in $\mathcal{R}$ is specified using one of the constraint relations in Table 1 applied to one or more nucleotides (e.g., specification of constraint $R^{\text{match}}_{a,b}$ requires that $\phi^a = \phi^b$ for nucleotides a and b).

In general, constraint satisfaction problems are NP-complete, so general-purpose polynomial-time algorithms are unavailable.[58] Empirically, we find that CSPs arising in the context of nucleic acid reaction pathway engineering specified in terms of the diverse constraint relations of Table 1 can typically be solved efficiently using the branch and propagate algorithm described below.

S1.11.2 Branch and Propagate Algorithm

We solve the CSP using a branch and propagate algorithm that returns a solution if one exists and returns a warning if no solution exists. Initially, the domain for each variable is $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{U}\}$. We first pre-process the CSP by trivially removing any value from the domain of a variable a that is inconsistent with a constraint (e.g., an assignment or library constraint). We further pre-process the CSP using constraint propagation to impose arc consistency as described below.

The branch and propagate algorithm involves iterated application of two ingredients:

- *constraint propagation* is used to narrow the search space by imposing *arc consistency* on each pair of variables: for any value in the domain of variable a there must be a consistent value in the domain of every other variable b, otherwise that value of variable a is inconsistent and can be removed from the domain of a (see "Chapter 3: Consistency-Enforcing and Constraint Propagation" of Reference 58).
- *depth-first branching* is used to extend a candidate partial solution by assigning a consistent value to one additional variable a, followed by *backtracking* to reassign the value of the most-recently assigned variable if no value in the domain of a is consistent with previous assignments (see "Chapter 5: General Search Strategies: Look-Ahead" of Reference 58).

S1.11.3 Feasible Sequence Inititialization

Sequence initialization (Section S1.7.1) commences with constraint propagation to impose arc consistency and then a first branching step in which a variable, a, is randomly selected and randomly assigned a value from the domain of a. Constraint propagation is then used to impose arc consistency and the next branching step is taken by randomly selecting an unassigned variable, b, and assigning a consistent value from the domain of b. Backtracking is performed if no consistent value for b exists. The branch and propagate algorithm returns a feasible set of initial sequences, $\phi_{\Psi^{\text{active}}}$, if one exists, and a warning otherwise.

S1.11.4 Feasible Sequence Mutation

During leaf mutation (Section S1.7.2), a feasible candidate sequence is generated by mutating the current leaf sequence, ϕ_{Λ_D}. This process begins with the sequence design algorithm randomly selecting a nucleotide a for mutation with probability (S12) and randomly assigning a new value from the domain of a. We then solve a CSP to obtain a valid candidate sequence consistent with the new value of a. Constraint propagation is used to impose arc consistency with the new value of a and any variables that require reassignment are added to the candidate mutation set, ξ_a. Initially, branching is performed by randomly selecting an unassigned variable b from ξ_a with probability proportional to the size of the domain of b (i.e., using weight

$$w_b = |\text{domain}(b)| \tag{S17}$$

to calculate the probability of selecting b). For each value ϕ^b in the domain of b, we check the implications of arc consistency on the size of the candidate mutation set, $\xi_{a,b}$, and create a priority queue based on the minimum increase in $|\xi_{a,b}|$ relative to $|\xi_a|$. Let $|\xi_{a,b_1}|$ denote the minimum increase, $|\xi_{a,b_2}|$ denote the next largest increase, and so on. Branching is performed by exploring the values of b according to their rank order in this priority queue. If no consistent value of b exists, backtracking is performed and the selection weight for variable b (S17) is updated using

$$w_b = \epsilon w_b + (1-\epsilon)(|\xi_{a,b_1}| - |\xi_a|). \tag{S18}$$

where $\epsilon = 0.5$ is a decay constant. The heuristics (S17) and (S18) seek to preferentially select highly constrained variables early in the branching process to avoid excessive backtracking. The initial weights (S17) assume that each variable b will imply a mutation set $\xi_{a,b}$ that increases in size with the size of domain(b), preferentially selecting variables with larger domains. With (S18), as we explore different variables b in ξ_a, we explicitly calculate the implied increase in $\xi_{a,b}$ due to each branching decision, and update the weights to bias future searching toward selection of b variables that cause the highest minimal increase in the size of the mutation set ξ_a.

The branch and propagate algorithm returns a feasible candidate sequence, $\hat{\phi}_{\Lambda_D}$, if one exists. Otherwise, the new value of a is invalid and is removed from the domain of a; a new value of a is randomly selected from the domain of a, and branch and propagate is applied again. If the only valid value of a is the current value, then m_{bad} is incremented and the leaf mutation procedure selects a new nucleotide for mutation with probability (S12).

S1.11.5 Feasible Sequence Reseeding

During leaf reoptimization (Section S1.7.3), a feasible candidate reseeded sequence, $\hat{\phi}_{\Lambda_D}$, is generated by introducing M_{reseed} feasible sequence mutations to the current leaf sequence, ϕ_{Λ_D}, via M_{reseed} consecutive calls to the

branch and propagate algorithm of Section S1.11.4 (selecting nucleotide a for mutation without replacement with probability (S12)).

S1.12 Pseudocode

OPTIMIZETUBES($\Omega, \Psi_\Omega^{\text{on}}, \Psi_\Omega^{\text{off}}, \Psi, s_\Psi, y_{\Omega,\Psi}, \mathcal{R}$)
  $\Psi^{\text{active}}, \Psi^{\text{passive}} \leftarrow \Psi^{\text{on}}, \Psi^{\text{off}}$
  $\phi_{\Psi^{\text{active}}} \leftarrow$ INITSEQ($s_{\Psi^{\text{active}}}, \mathcal{R}$)
  $\Lambda, D \leftarrow$ MAKEFOREST($s_{\Psi^{\text{active}}}$)
  $\phi_\Lambda, \tilde{\mathcal{M}}_1 \leftarrow$ OPTIMIZEFOREST(ϕ_Λ, D)
  $\mathcal{M} \leftarrow$ EVALUATEDEFECT(ϕ_Ψ)
  $\hat{\phi}_\Psi, \hat{\mathcal{M}} \leftarrow \phi_\Psi, \mathcal{M}$
  while $\hat{\mathcal{M}} > \max(f_{\text{stop}}, \tilde{\mathcal{M}}_1)$
    $\Psi^{\text{active}}, \Psi^{\text{passive}} \leftarrow$ REFOCUSTUBES($\Psi^{\text{active}}, \Psi^{\text{passive}}, \{\hat{x}_{h,\Psi_h^{\text{passive}}}\}$)
    $\Lambda, D \leftarrow$ AUGMENTFOREST($\Lambda, D, \hat{P}_{\Psi^{\text{active}}}$)
    $\hat{\phi}_\Lambda, \tilde{\mathcal{M}}_1 \leftarrow$ OPTIMIZEFOREST($\hat{\phi}_\Lambda, D$)
    $\hat{\mathcal{M}} \leftarrow$ EVALUATEDEFECT($\hat{\phi}_\Psi$)
    if $\hat{\mathcal{M}} < \mathcal{M}$
      $\phi_\Psi, \mathcal{M} \leftarrow \hat{\phi}_\Psi, \hat{\mathcal{M}}$
  return ϕ_Ψ

OPTIMIZEFOREST(ϕ_Λ, D)
  $\tilde{\mathcal{M}}_d \leftarrow \infty \ \forall d \in \{1, \ldots, D\}$
  $\beta_{\text{merge}} \leftarrow$ **false**
  while $\neg\beta_{\text{merge}}$
    $\phi_{\Lambda_D}, \tilde{\mathcal{M}}_D \leftarrow$ OPTIMIZELEAVES(ϕ_{Λ_D}, D)
    $d \leftarrow D - 1$
    $\beta_{\text{merge}} \leftarrow$ **true**
    while $d \geq 1$ **and** β_{merge}
      $\hat{\phi}_{\Lambda_d} \leftarrow$ MERGESEQ($\phi_{\Lambda_{d+1}}$)
      $\hat{\mathcal{M}}_d \leftarrow$ ESTIMATEDEFECT($\hat{\phi}_{\Lambda_d}$)
      if $\hat{\mathcal{M}}_d < \tilde{\mathcal{M}}_d$
        $\phi_{\Lambda_d}, \tilde{\mathcal{M}}_d \leftarrow \hat{\phi}_{\Lambda_d}, \hat{\mathcal{M}}_d$
      if $\hat{\mathcal{M}}_d > \max(f_d^{\text{stop}}, \tilde{\mathcal{M}}_{d+1}/f_{\text{stringent}})$
        $\beta_{\text{merge}} \leftarrow$ **false**
        $\Lambda, D \leftarrow$ REDECOMPOSEFOREST($\Lambda, D, s_{\Lambda_d}, \hat{P}_{\Lambda_d}$)
        $\phi_{\Lambda_D} \leftarrow$ SPLITSEQ($\hat{\phi}_{\Lambda_d}$)
        $\tilde{\mathcal{M}}_{d'} \leftarrow \infty \ \forall d' \in \{d+1, \ldots, D\}$
      $d \leftarrow d - 1$
  return $\phi_{\Lambda_1}, \tilde{\mathcal{M}}_1$

OPTIMIZELEAVES(ϕ_{Λ_D}, D)
  $\phi_{\Lambda_D}, \tilde{\mathcal{M}}_D \leftarrow$ MUTATELEAVES(ϕ_{Λ_D}, D)
  $m_{\text{reopt}} \leftarrow 0$
  while $\tilde{\mathcal{M}}_D > f_D^{\text{stop}}$ **and** $m_{\text{reopt}} < M_{\text{reopt}}$
    $\hat{\phi}_{\Lambda_D} \leftarrow$ RESEEDSEQ($\phi_{\Lambda_D}, \{\tilde{\mathcal{M}}_{h,j}^a\}, \mathcal{R}$)
    $\hat{\phi}_{\Lambda_D}, \hat{\mathcal{M}}_D \leftarrow$ MUTATELEAVES($\hat{\phi}_{\Lambda_D}, D$)
    if $\hat{\mathcal{M}}_D < \tilde{\mathcal{M}}_D$
      $\phi_{\Lambda_D}, \tilde{\mathcal{M}}_D \leftarrow \hat{\phi}_{\Lambda_D}, \hat{\mathcal{M}}_D$
      $m_{\text{reopt}} \leftarrow 0$
    else
      $m_{\text{reopt}} \leftarrow m_{\text{reopt}} + 1$
  return $\phi_{\Lambda_D}, \tilde{\mathcal{M}}_D$

MUTATELEAVES(ϕ_{Λ_D}, D)
  $\tilde{\mathcal{M}}_D \leftarrow$ ESTIMATEDEFECT(ϕ_{Λ_D})
  $\gamma_{\text{bad}} \leftarrow \emptyset, \quad m_{\text{bad}} \leftarrow 0$
  while $\tilde{\mathcal{M}}_D > f_D^{\text{stop}}$ **and** $m_{\text{bad}} < M_{\text{bad}}$
    $\xi, \hat{\phi}_{\Lambda_D} \leftarrow$ SAMPLEMUTATION($\phi_{\Lambda_D}, \{\tilde{\mathcal{M}}_{h,j}^a\}, \mathcal{R}$)
    if $\xi \in \gamma_{\text{bad}}$
      $m_{\text{bad}} \leftarrow m_{\text{bad}} + 1$
    else
      $\hat{\mathcal{M}}_D \leftarrow$ ESTIMATEDEFECT($\hat{\phi}_{\Lambda_D}$)
      if $\hat{\mathcal{M}}_D < \tilde{\mathcal{M}}_D$
        $\phi_{\Lambda_D}, \tilde{\mathcal{M}}_D \leftarrow \hat{\phi}_{\Lambda_D}, \hat{\mathcal{M}}_D$
        $\gamma_{\text{bad}} \leftarrow \emptyset, \quad m_{\text{bad}} \leftarrow 0$
      else
        $\gamma_{\text{bad}} \leftarrow \gamma_{\text{bad}} \cup \xi, \quad m_{\text{bad}} \leftarrow m_{\text{bad}} + 1$
  return $\phi_{\Lambda_D}, \tilde{\mathcal{M}}_D$

ESTIMATEDEFECT(ϕ_{Λ_d})
  $\tilde{Q}_{\Lambda_d}, \tilde{P}_{\Lambda_d} \leftarrow$ CONDITIONALNODALPROPERTIES(ϕ_{Λ_d})
  $\tilde{Q}_{\Psi^{\text{active}}} \leftarrow$ ESTIMATECOMPLEXPFUNCS($\tilde{Q}_{\Lambda_d}$)
  $\tilde{P}_{\Psi^{\text{active}}} \leftarrow$ ESTIMATECOMPLEXPAIRPROBS($\tilde{P}_{\Lambda_d}$)
  for $h \in \Omega$
    $\tilde{x}^0_{h,\Psi_h^0} \leftarrow$ DEFLATEMASSCONSTRAINTS(x^0_{h,Ψ_h^0})
    $\tilde{x}_{h,\Psi_h^{\text{active}}} \leftarrow$ ESTIMATECOMPLEXCONCS($\tilde{Q}_{\Psi_h^{\text{active}}}, \tilde{x}^0_{h,\Psi_h^0}$)
    $\{\tilde{\mathcal{M}}_{h,j}^a\} \leftarrow$ ESTIMATECONTRIBS($\tilde{P}_{\Psi_h^{\text{on}}}, s_{\Psi_h^{\text{on}}}, \tilde{x}_{h,\Psi_h^{\text{on}}}, y_{h,\Psi_h^{\text{on}}}$)
  $\tilde{\mathcal{M}}_d \leftarrow \sum_{h\in\Omega} \sum_{j\in\Psi_h^{\text{on}}} \sum_{1\leq a\leq|\phi_j|} \tilde{\mathcal{M}}_{h,j}^a$
  return $\tilde{\mathcal{M}}_d$

Algorithm S1. Pseudocode for constrained multistate test tube ensemble defect optimization. Consider the set of target test tubes, Ω, collectively containing the set of complexes Ψ (comprising the sets of on-target complexes Ψ_Ω^{on} and off-target complexes Ψ_Ω^{off}) with target secondary structures s_Ψ and target concentrations $y_{\Omega,\Psi}$. The function call OPTIMIZETUBES($\Omega, \Psi_\Omega^{\text{on}}, \Psi_\Omega^{\text{off}}, \Psi, s_\Psi, y_{\Omega,\Psi}, \mathcal{R}$) returns the set of designed sequences, ϕ_Ψ, satisfying the sequence constraints in $\mathcal{R}$.

S1.13 Default Algorithm Parameters

Default algorithm parameters are shown in Table S2.

Table S2. RNA design: default algorithm parameters for constrained multistate test tube ensemble defect optimization.

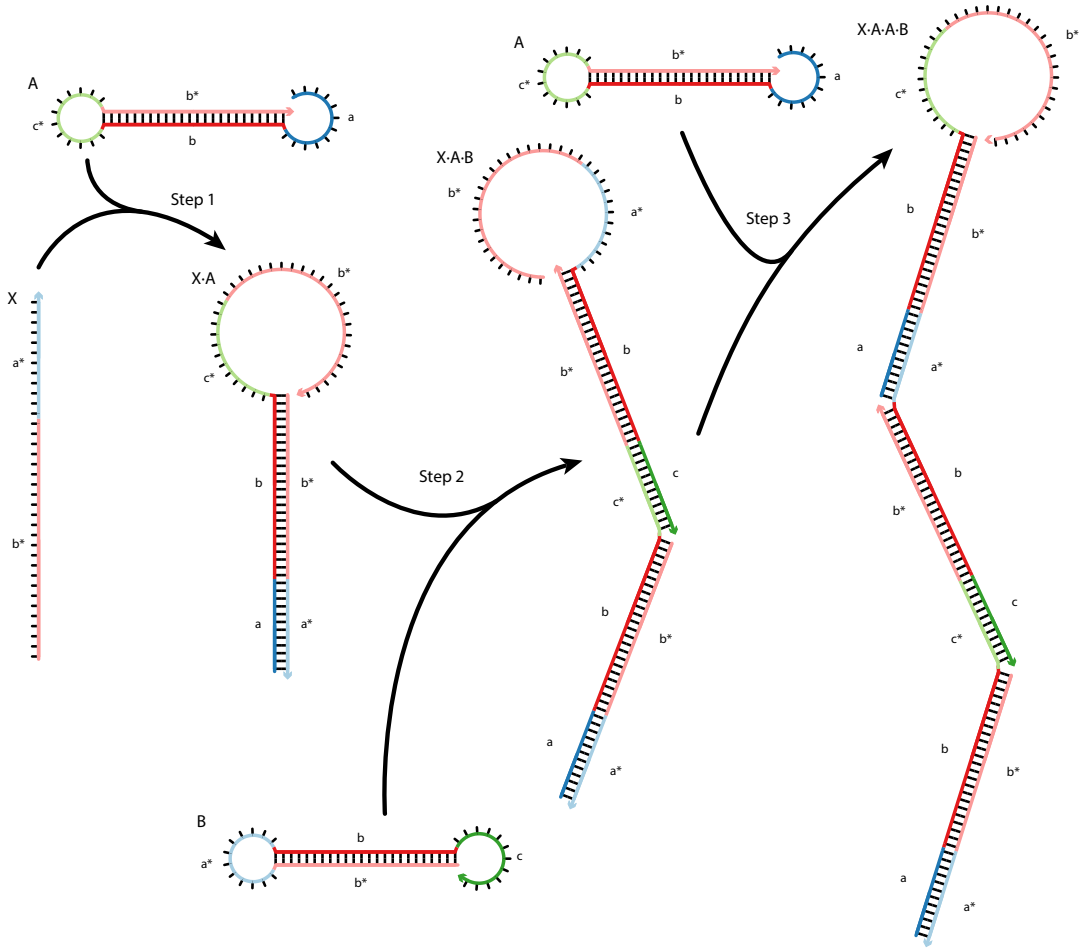
Parameter	Value
f_{stop}	0.02
f_{passive}	0.01
H_{split}	2
N_{split}	12
f_{split}	0.99
$f_{\text{stringent}}$	0.99
ΔG^{clamp}	-25 kcal/mol
M_{bad}	300
M_{reseed}	50
M_{reopt}	3
f_{redecomp}	0.03
f_{refocus}	0.03

For DNA design, $H_{\text{split}} = 3$.

S2 Engineering Case Studies

S2.1 Reaction Pathways

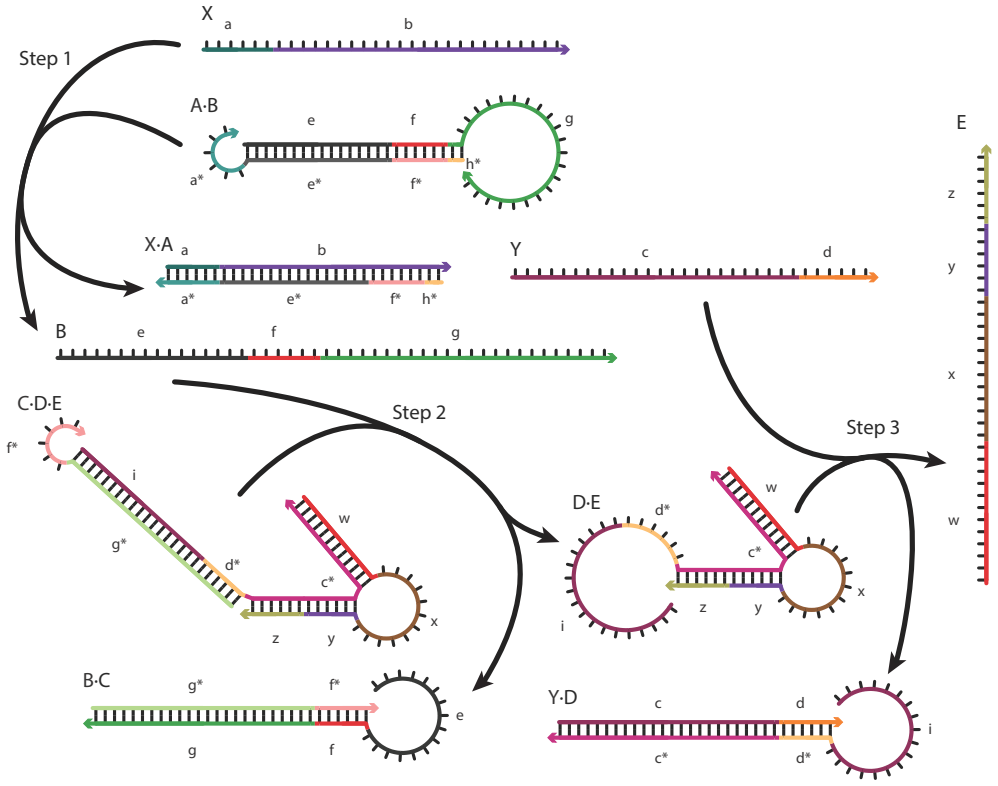
S2.1.1 Conditional Self-Assembly via Hybridization Chain Reaction (HCR)



Step	Reaction	Function	Mechanism
1	$X + A \rightarrow X{\cdot}A$	detect target, first A polymerization step	toehold/toehold nucleation, 3-way branch migration
2	$X{\cdot}A + B \rightarrow X{\cdot}A{\cdot}B$	first B polymerization step, regenerate target sequence	toehold/toehold nucleation, 3-way branch migration
$2k$+1	$X{\cdot}(A)_k{\cdot}(B)_k + A \rightarrow X{\cdot}(A)_{k+1}{\cdot}(B)_k$	generic A polymerization step, $k = 1, 2, \ldots$	toehold/toehold nucleation, 3-way branch migration
$2k$+2	$X{\cdot}(A)_{k+1}{\cdot}(B)_k + B \rightarrow X{\cdot}(A)_{k+1}{\cdot}(B)_{k+1}$	generic B polymerization step, $k = 1, 2, \ldots$	toehold/toehold nucleation, 3-way branch migration

Figure S1. Reaction pathway for conditional self-assembly via hybridization chain reaction (HCR).[20] Target X triggers self-assembly of metastable hairpins A and B into a long nicked dsDNA polymer via a chain reaction of alternating A and B polymerization steps. Top: Reaction pathway schematic. Bottom: Elementary step details.

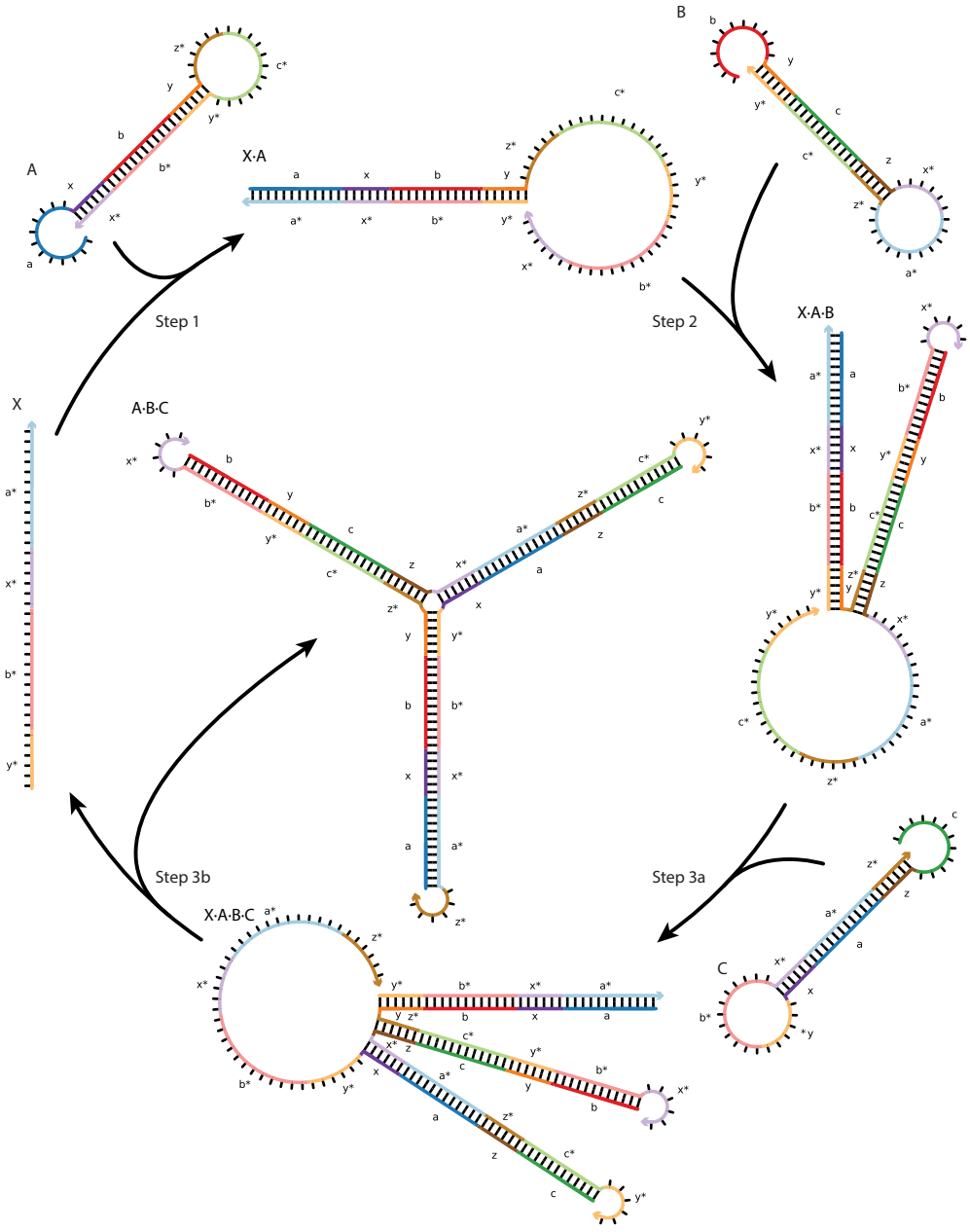
S2.1.2 Boolean Logic AND using Toehold Sequestration Gates



Step	Reaction	Function	Mechanism
1	X + A·B → X·A + B	translate X target sequence	toehold/toehold nucleation, 3-way branch migration
2	B + C·D·E → B·C + D·E	detect translated first target	toehold/toehold nucleation, 3-way branch migration, expose toehold
3	Y + D·E → Y·D + E	detect second target	toehold/toehold nucleation, 3-way branch migration

Figure S2. Reaction pathway for Boolean logic AND using toehold sequestration gates.[59] Gates implement the logical operation "if targets X AND Y are detected, generate output E". Top: Reaction pathway schematic. Bottom: Elementary step details.

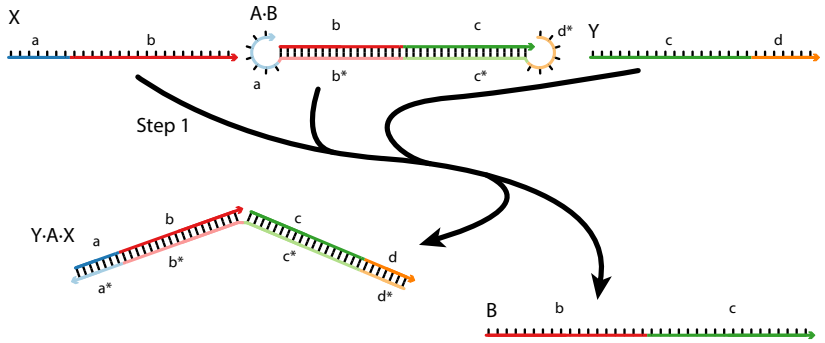
S2.1.3 Self-Assembly of a 3-Arm Junction via Catalytic Hairpin Assembly (CHA)



Step	Reaction	Function	Mechanism
1	X + A → X·A	assemble with catalyst X	toehold/toehold nucleation, 3-way branch migration
2	X·A + B → X·A·B	assemble	toehold/toehold nucleation, 3-way branch migration
3a	X·A·B + C → X·A·B·C	assemble	toehold/toehold nucleation, 3-way branch migration
3b	X·A·B·C → X + A·B·C	disassemble from catalyst X and assemble 3-arm junction	intracomplex blunt-end strand invasion, 3-way branch migration

Figure S3. Reaction pathway for self-assembly of a 3-arm junction via catalytic hairpin assembly (CHA).[24] Target X catalyzes self-assembly of metastable hairpins A, B, and C into 3-arm junction A·B·C. Top: Reaction pathway schematic. Bottom: Elementary step details.

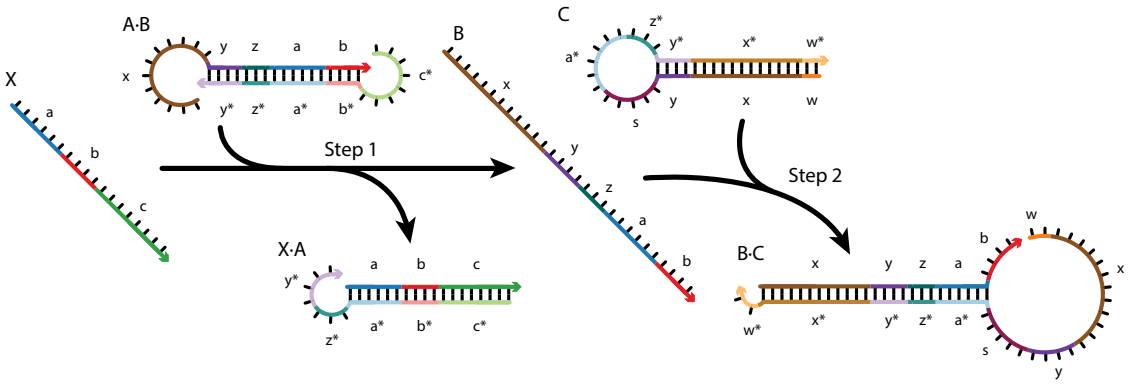
S2.1.4 Boolean Logic AND using a Cooperative Hybridization Gate



Step	Reaction	Function	Mechanism
1	X + Y + A·B → Y·A·X + B	cooperative detection of X and Y to generate output B	toehold/toehold nucleation, 3-way branch migration

Figure S4. Reaction pathway for Boolean logic AND using a cooperative hybridization gate.[60] Gate implements the logical operation "if targets X AND Y are detected, cooperatively generate output B". Top: Reaction pathway schematic. Bottom: Elementary step details.

S2.1.5 Conditional Dicer Substrate Formation via Shape and Sequence Transduction with Small Conditional RNAs (scRNAs)



Step	Reaction	Function	Mechanism
1	X + A·B → X·A + B	detect target X (sequence 'a-b-c')	toehold/toehold nucleation, 3-way branch migration, spontaneous dissociation
2	B + C → B·C	form Dicer substrate targeting independent target Y (sequence 'w-x-y-z')	toehold/loop nucleation, 3-way branch migration

Figure S5. Reaction pathway for conditional Dicer substrate formation via shape and sequence transduction with small conditional RNAs (scRNAs).[33] scRNA A·B detects target X (comprising sequence 'a-b-c'), generating intermediate B that assembles with scRNA C to generate Dicer substrate B·C (targeting independent sequence 'w-x-y-z' for silencing). Top: Reaction pathway schematic. Bottom: Elementary step details.

S2.2 Specification of Target Test Tubes

S2.2.1 General Formulation

Consider specification of the multistate test tube ensemble, Ω, for the design of N orthogonal systems for a reaction pathway of M elementary steps, each corresponding to a self-assembly or disassembly operation in which complexes form or break. One elementary step tube is specified for each step $m = 0, 1, \ldots, M$ for each system $n = 1, \ldots, N$ (treating formation of the initial reactants as a precursor "Step 0"). Additionally, a single global crosstalk tube is specified to minimize off-pathway interactions between the reactive species generated during all elementary steps of all systems. The total number of target test tubes is then $|\Omega| = (M + 1) \times N + 1$.

Elementary Step Tubes. Consider elementary step m for orthogonal system n with on-pathway products $\Psi_{m_n}^{\text{products}}$ that are intended to form at non-zero concentrations at equilibrium, and reactants $\Psi_{m_n}^{\text{reactants}}$ that are intended to fully convert into the on-pathway products at equilibrium. Furthermore, consider the set of off-pathway products, $\Psi_{m_n}^{\text{crosstalk}}$, corresponding to unintended interactions between these same reactants.

The *elementary step tube* for step m of system n is then:

$$\boxed{\text{Step } m_n \text{ tube: } \Psi_h^{\text{on}} \equiv \Psi_{m_n}^{\text{products}}, \quad \Psi_h^{\text{off}} \equiv \Psi_{m_n}^{\text{reactants}} \cup \Psi_{m_n}^{\text{crosstalk}}}$$

where the on-targets are the on-pathway products, and the off-targets are the reactants and off-pathway crosstalk products. For step m of system n, this tube designs for full conversion of cognate reactants into cognate products and against local crosstalk between these same reactants. One elementary step tube is specified for each elementary step $m = 0, 1, \ldots, M$ for each system $n = 1, \ldots, N$.

The off-pathway crosstalk products for step m of system n are defined as:

$$\boxed{\Psi_{m_n}^{\text{crosstalk}} = \Psi_{m_n}^{L \leq L_{\max}} - \Psi_{m_n}^{\text{exclude}}}$$

where the set $\Psi_{m_n}^{L \leq L_{\max}}$ denotes the set of all complexes of up to $L_{\max}$ strands (that are not already on-targets in the Step m_n tube). The set $\Psi_{m_n}^{\text{exclude}}$ contains energetically favorable complexes that we wish to exclude from the ensemble for the current elementary step (e.g., downstream on-pathway products, or off-pathway products that are inhibited kinetically rather than thermodynamically, and hence are not suitable for inclusion in the equilibrium optimization ensemble).

Global Crosstalk Tube. To actively design against global crosstalk, we additionally specify a single *global crosstalk tube*:

$$\boxed{\text{Global crosstalk tube: } \Psi_h^{\text{on}} \equiv \Psi_{\text{global}}^{\text{reactive}}, \quad \Psi_h^{\text{off}} \equiv \Psi_{\text{global}}^{\text{crosstalk}}}$$

where $\Psi_{\text{global}}^{\text{reactive}}$ denotes the set of all reactive species generated during all elementary steps for all systems and $\Psi_{\text{global}}^{\text{crosstalk}}$ denotes the set of undesired crosstalk products resulting from interactions between these species.

For the global crosstalk tube, we exploit *motif simplification* to enable specification of the on-target and off-target complexes using using only monomers and dimers. The presumption is that motif complexity will typically decrease rather than increase crosstalk between reactive species, so that for the global crosstalk tube, motif simplification is justified in the service of efficiency and simplicity. By contrast, for the elementary step tubes, reactant and product complexes are treated without motif simplification, ensuring that any energetic effects associated with the full complexes (either unfavorable [e.g., 3-arm junction for CHA product] or favorable [e.g., nick stack for HCR]) are taken into consideration.

To define various forms of motif simplification, it is helpful to define input and output domains that participate in the elementary steps. Each scRNA or scDNA motif (monomer, dimer, trimer, etc) has one or more *input domains* that control the state of one or more *output domains*. An inactive output domain is toggled to the active state when

sequestering input domains hybridize to active output domains generated by earlier elementary steps in the reaction pathway. Nucleation with an input domain occurs via hybridization to an accessible loop or toehold. Targets that serve as inputs to a reaction pathway may be viewed as unconditionally active output domains that are available to hybridize to complementary input domains at any step in a reaction pathway.

Using motif simplification, we specify the *reactive species* and *cognate products* for system n as follows:

- $\lambda_n^{\text{simple}}$: scRNA and scDNA motifs with multiple input or output domains are simplified so that only the input and output domains for a single elementary step are present in each simplified motif.
- $\lambda_n^{\text{ss-out}}$: single-stranded output domains are specified for each elementary step, removing other concatenated or hybridized domains that represent the history or future of the reaction (participating in previous or future elementary steps).
- $\lambda_n^{\text{ss-in}}$: single-stranded nucleation sites within input domains (toeholds or loops) are specified isolated from the surrounding domains representing the history or future of the reaction.*
- $\lambda_n^{\text{reactive}} \equiv \{\lambda^{\text{simple}} \cup_{\text{simp}} \lambda^{\text{ss-out}} \cup_{\text{simp}} \lambda^{\text{ss-in}}\}_n$: the set of reactive species for system n is specified using a union operator $\cup_{\text{simp}}$ that eliminates redundancies when one monomer species is an accessible subsequence of another monomer species.
- $\lambda_n^{\text{cognate}}$: cognate products expected to form from reactive species in $\lambda_n^{\text{reactive}}$ based on sequence complementarity imposed by the reaction pathway (e.g., an input domain within a motif in $\lambda_n^{\text{simple}}$ is expected to hybridize to a complementary output domain in $\lambda_n^{\text{ss-out}}$).

These definitions for $\lambda_n^{\text{reactive}}$ and $\lambda_n^{\text{cognate}}$ are then used to define the on-targets for the global crosstalk tube:

$$\boxed{\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}}$$

and the off-targets for the global crosstalk tube:

$$\boxed{\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}}$$

Here, $\Psi_{\text{global}}^{L\leq L_{\max}}$ denotes the set of all complexes of up to $L_{\max}$ strands (that are not already on-targets in the global crosstalk tube). The set $\cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$ contains all the cognate products that the reactive species in the N orthogonal systems are expected to form based on sequence complementarity. Crucially, by excluding these cognate products from $\Psi_{\text{global}}^{\text{crosstalk}}$, they do not appear in the global crosstalk tube as either on-targets or off-targets. Hence, all reactive species in the global crosstalk tube are forced to perform either no reaction (remaining as desired on-targets) or to undergo a crosstalk reaction (forming undesired off-targets), providing the basis for minimization of global crosstalk during sequence optimization.

*The role of $\lambda_n^{\text{ss-in}}$ is to enable $\Psi_{\text{global}}^{\text{crosstalk}}$ (the off-targets for the global crosstalk tube) to be specified without requiring any complexes larger than dimers. For example, consider a dimer motif with an exposed toehold. Crosstalk via kissing of this toehold with that of another dimer motif would yield a tetramer off-target; including these toeholds as isolated monomers in $\lambda_n^{\text{ss-in}}$ allows this crosstalk interaction to be described by an off-target dimer, which is automatically included in $\Psi_{\text{global}}^{\text{crosstalk}}$ by considering all off-targets of up to $L_{\max} = 2$ strands. Further, inclusion of loop nucleation sites in $\lambda_n^{\text{ss-in}}$ enables designing against pseudoknotted toehold/loop and loop/loop crosstalk interactions without needing to explicitly include pseudoknots in the structural ensemble of any complex.

S2.2.2 Conditional Self-Assembly via HCR

Target test tubes are defined using the specification of Section S2.2.1 with the following definitions. The total number of target test tubes is $|\Omega| = \sum_{n=1,\ldots,N}\{\text{Step 0, Step 1, Step 2, Step 3}\}_n + \text{Crosstalk} = 4N + 1$; the target test tubes in the multistate test tube ensemble, Ω, are indexed by $h = 1, \ldots, 4N + 1$. $L_{\max} = 2$ for all tubes.

Reactants for system n

- Target: X_n
- Hairpins: $\{\text{A, B}\}_n$

Elementary step tubes for system *n*

- Step 0_n tube: $\Psi_{0_n}^{\text{products}} \equiv \{\text{X, A, B}\}_n$; $\Psi_{0_n}^{\text{reactants}} \equiv \{\text{A·B}\}_n$ (dimer nucleus that inhibits leakage); $\Psi_{0_n}^{\text{exclude}} \equiv \{\text{X·A}\}_n$ (downstream on-pathway product)
- Step 1_n tube: $\Psi_{1_n}^{\text{products}} \equiv \{\text{X·A}\}_n$; $\Psi_{1_n}^{\text{reactants}} \equiv \{\text{X, A}\}_n$; $\Psi_{1_n}^{\text{exclude}} \equiv \emptyset$
- Step 2_n tube: $\Psi_{2_n}^{\text{products}} \equiv \{\text{X·A·B}\}_n$; $\Psi_{2_n}^{\text{reactants}} \equiv \{\text{X·A, B}\}_n$; $\Psi_{2_n}^{\text{exclude}} \equiv \emptyset$
- Step 3_n tube: $\Psi_{3_n}^{\text{products}} \equiv \{\text{X·A·A·B}\}_n$; $\Psi_{3_n}^{\text{reactants}} \equiv \{\text{X·A·B, A}\}_n$; $\Psi_{3_n}^{\text{exclude}} \equiv \emptyset$

Global crosstalk tube

- Crosstalk tube: $\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$; $\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L \le L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

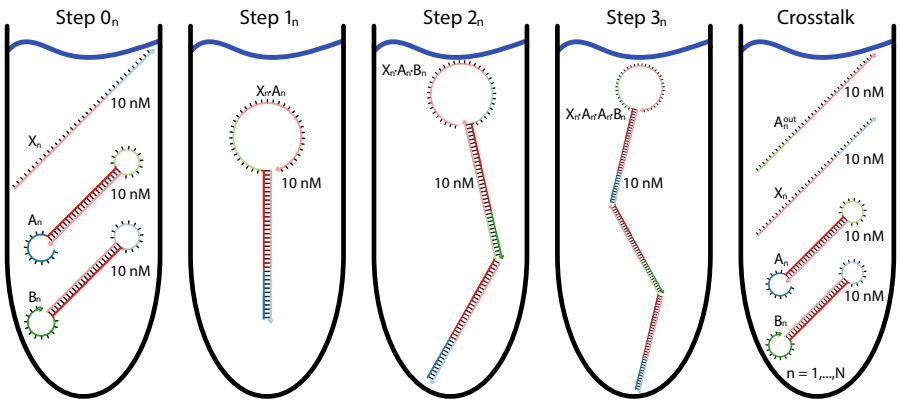
The reactive species and cognate products for system n are:

- $\lambda_n^{\text{simple}} \equiv \{\text{A, B}\}_n$
- $\lambda_n^{\text{ss-out}} \equiv \{\text{X, A}^{\text{out}}, \text{B}^{\text{out}}\}_n$
- $\lambda_n^{\text{ss-in}} \equiv \{\text{A}^{\text{toe}}, \text{B}^{\text{toe}}\}_n$
- $\lambda_n^{\text{reactive}} \equiv \{\text{A, B, A}^{\text{out}}, \text{B}^{\text{out}}\}_n$
- $\lambda_n^{\text{cognate}} \equiv \{\text{A}^{\text{out}}\text{·B, B}^{\text{out}}\text{·A}\}_n$

based on the definitions (listed 5′ to 3′ using the sequence domain notation of Figure S1):

- $\text{A} \equiv \text{A}^{\text{in}}\text{-A}^{\text{out}}$
- $\text{A}^{\text{toe}} \equiv \text{a}$
- $\text{A}^{\text{in}} \equiv \text{a-b}$
- $\text{A}^{\text{out}} \equiv \text{c*-b*}$
- $\text{B} \equiv \text{B}^{\text{out}}\text{-B}^{\text{in}}$
- $\text{B}^{\text{toe}} \equiv \text{c}$
- $\text{B}^{\text{in}} \equiv \text{b-c}$
- $\text{B}^{\text{out}} \equiv \text{b*-a*}$
- $\text{X} \equiv \text{b*-a*}$

Note: X_n is identical to B_n^{out}, so it is implicitly included in the definition of $\lambda_n^{\text{reactive}}$. To avoid redundancy, the toeholds of $\lambda_n^{\text{ss-in}}$ are not included in the definition of $\lambda_n^{\text{reactive}}$; these toeholds are already available to form dimer crosstalk products in the hairpin monomers of $\lambda_n^{\text{simple}}$.



Tube	On-targets (Ψ_h^{on})	Off-targets (Ψ_h^{off})
Step 0_n	$\{\text{X, A, B}\}_n$	$\{\text{A·B}\}_n \cup \Psi_{0_n}^{L\leq L_{\max}} - \{\text{X·A}\}_n$
Step 1_n	$\{\text{X·A}\}_n$	$\{\text{X, A}\}_n \cup \Psi_{1_n}^{L\leq L_{\max}}$
Step 2_n	$\{\text{X·A·B}\}_n$	$\{\text{X·A, B}\}_n \cup \Psi_{2_n}^{L\leq L_{\max}}$
Step 3_n	$\{\text{X·A·A·B}\}_n$	$\{\text{X·A·B, A}\}_n \cup \Psi_{3_n}^{L\leq L_{\max}}$
Crosstalk	$\cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$	$\Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

Figure S6. Target test tubes for conditional self-assembly via HCR (reaction pathway of Figure S1). Top: Target test tube schematics. Bottom: Target test tube details. Each target test tube contains the depicted on-target complexes (each with the depicted target structure and a target concentration of 10 nM) and the off-target complexes listed in the table (each with vanishing target concentration). To simultaneously design N orthogonal systems, the total number of target test tubes is $|\Omega| = 4N + 1$. $L_{\max} = 2$ for all tubes. Design conditions: DNA in 1 M Na^+ at 25 °C.

S2.2.3 Boolean Logic AND using Toehold Sequestration Gates

Target test tubes are defined using the specification of Section S2.2.1 with the following definitions. The total number of target test tubes is $|\Omega| = \sum_{n=1,\ldots,N} \{\text{Step 0, Step 1, Step 2, Step 3}\}_n$ + Crosstalk = $4N + 1$; the target test tubes in the multistate test tube ensemble, Ω, are indexed by $h = 1, \ldots, 4N + 1$. $L_{\max} = 2$ for all tubes.

Reactants for system n

- Targets: $\{\text{X, Y}\}_n$
- Translator gate: $\{\text{A·B}\}_n$
- AND gate: $\{\text{C·D·E}\}_n$

Elementary step tubes for system n

- Step 0_n: $\Psi_{0_n}^{\text{products}} \equiv \{\text{X, Y, A·B, C·D·E}\}_n$; $\Psi_{0_n}^{\text{reactants}} \equiv \{\text{A, B, C, D, E, C·D, D·E}\}_n$; $\Psi_{0_n}^{\text{exclude}} \equiv \{\text{X·A}\}_n$
- Step 1_n: $\Psi_{1_n}^{\text{products}} \equiv \{\text{X·A, B}\}_n$; $\Psi_{1_n}^{\text{reactants}} \equiv \{\text{X, A·B}\}_n$; $\Psi_{1_n}^{\text{exclude}} \equiv \emptyset$
- Step 2_n: $\Psi_{2_n}^{\text{products}} \equiv \{\text{B·C, D·E}\}_n$; $\Psi_{2_n}^{\text{reactants}} \equiv \{\text{B, C·D·E}\}_n$; $\Psi_{2_n}^{\text{exclude}} \equiv \emptyset$
- Step 3_n: $\Psi_{3_n}^{\text{products}} \equiv \{\text{Y·D, E}\}_n$; $\Psi_{3_n}^{\text{reactants}} \equiv \{\text{Y, D·E}\}_n$; $\Psi_{3_n}^{\text{exclude}} \equiv \emptyset$

Global crosstalk tube

- Crosstalk tube: $\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$; $\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L \le L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

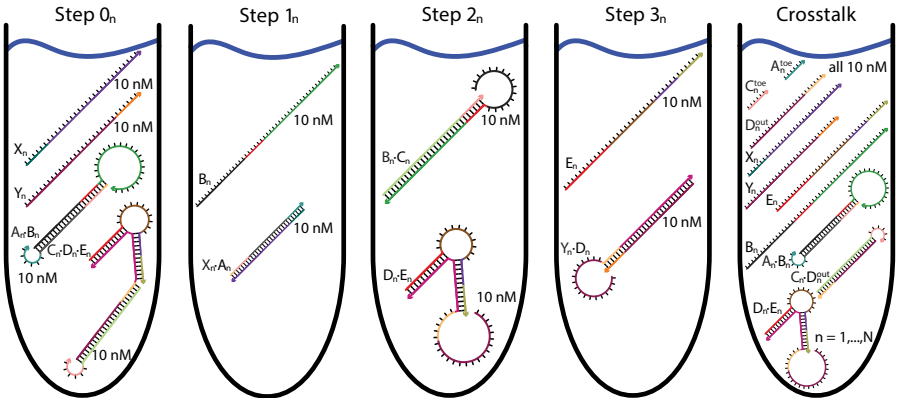
The reactive species and cognate products for system n are:

- $\lambda_n^{\text{simple}} \equiv \{\text{A·B, C·D}^{\text{out}}\text{, D·E}\}_n$
- $\lambda_n^{\text{ss-out}} \equiv \{\text{X, Y, B, D}^{\text{out}}\text{, E}\}_n$
- $\lambda_n^{\text{ss-in}} \equiv \{\text{A}^{\text{toe}}\text{, C}^{\text{toe}}\text{, D}^{\text{toe}}\}_n$
- $\lambda_n^{\text{reactive}} \equiv \{\text{A·B, C·D}^{\text{out}}\text{, D·E, X, Y, B, D}^{\text{out}}\text{, E, A}^{\text{toe}}\text{, C}^{\text{toe}}\}_n$
- $\lambda_n^{\text{cognate}} \equiv \{\text{X·A, B·C, Y·D, X·A}^{\text{toe}}\text{, B·C}^{\text{toe}}\}_n$

based on the definitions (listed $5'$ to $3'$ using the sequence domain notation of Figure S2):

- $\text{A} \equiv \text{h*-f*-e*-a*}$
- $\text{A}^{\text{toe}} \equiv \text{a*}$
- $\text{B} \equiv \text{e-f-g}$
- $\text{C} \equiv \text{g*-f*}$
- $\text{C}^{\text{toe}} \equiv \text{f*}$
- $\text{D} \equiv \text{D}^{\text{out}}\text{-D}^{\text{in}}$
- $\text{D}^{\text{toe}} \equiv \text{d*}$
- $\text{D}^{\text{in}} \equiv \text{c*}$
- $\text{D}^{\text{out}} \equiv \text{i-d*}$
- $\text{E} \equiv \text{w-x-y-z}$
- $\text{X} \equiv \text{a-b}$
- $\text{Y} \equiv \text{c-d}$

Note: D_n^{toe} is contained in D_n^{out}, providing a toehold adjacent to D_n^{in}. To avoid redundancy, we omit D_n^{toe} from $\lambda_n^{\text{reactive}}$ because it is a subsequence of D_n^{out} in $\lambda_n^{\text{ss-out}}$.



Tube	On-targets (Ψ_h^{on})	Off-targets (Ψ_h^{off})
Step 0_n	$\{\text{X, Y, A·B, C·D·E}\}_n$	$\{\text{A, B, C, D, E, C·D, D·E}\}_n \cup \Psi_{0_n}^{L\leq L_{\max}} - \{\text{X·A}\}_n$
Step 1_n	$\{\text{X·A, B}\}_n$	$\{\text{X, A·B}\}_n \cup \Psi_{1_n}^{L\leq L_{\max}}$
Step 2_n	$\{\text{B·C, D·E}\}_n$	$\{\text{B, C·D·E}\}_n \cup \Psi_{2_n}^{L\leq L_{\max}}$
Step 3_n	$\{\text{Y·D, E}\}_n$	$\{\text{Y, D·E}\}_n \cup \Psi_{3_n}^{L\leq L_{\max}}$
Crosstalk	$\cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$	$\Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

Figure S7. Target test tubes for Boolean logic AND using toehold sequestration gates (reaction pathway of Figure S2). Top: Target test tube schematics. Bottom: Target test tube details. Each target test tube contains the depicted on-target complexes (each with the depicted target structure and a target concentration of 10 nM) and the off-target complexes listed in the table (each with vanishing target concentration). To simultaneously design N orthogonal systems, the total number of target test tubes is $|\Omega| = 4N + 1$. $L_{\max} = 2$ for all tubes. Design conditions: DNA in 1 M Na^+ at 25 °C.

S2.2.4 Self-Assembly of a 3-Arm Junction via CHA

Target test tubes are defined using the specification of Section S2.2.1 with the following definitions. The total number of target test tubes is $|\Omega| = \sum_{n=1,\ldots,N} \{\text{Step 0, Step 1, Step 2, Step 3a, Step 3b}\}_n$ + Crosstalk = $5N + 1$; the target test tubes in the multistate test tube ensemble, Ω, are indexed by $h = 1, \ldots, 5N + 1$. $L_{\max} = 2$ for all tubes.

Reactants for system n

- Target: X_n
- Hairpins: $\{\mathrm{A, B, C}\}_n$

Elementary step tubes for system n

- Step 0_n: $\Psi_{0_n}^{\text{products}} \equiv \{\mathrm{X, A, B, C}\}_n$; $\Psi_{0_n}^{\text{reactants}} \equiv \emptyset$; $\Psi_{0_n}^{\text{exclude}} \equiv \{\mathrm{X{\cdot}A}\}_n$
- Step 1_n: $\Psi_{1_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A}\}_n$; $\Psi_{1_n}^{\text{reactants}} \equiv \{\mathrm{X, A}\}_n$; $\Psi_{1_n}^{\text{exclude}} \equiv \emptyset$
- Step 2_n: $\Psi_{2_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A{\cdot}B}\}_n$; $\Psi_{2_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A, B}\}_n$; $\Psi_{2_n}^{\text{exclude}} \equiv \emptyset$
- Step $3a_n$: $\Psi_{3a_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A{\cdot}B{\cdot}C}\}_n$; $\Psi_{3a_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A{\cdot}B, C}\}_n$; $\Psi_{3a_n}^{\text{exclude}} \equiv \emptyset$
- Step $3b_n$: $\Psi_{3b_n}^{\text{products}} \equiv \{\mathrm{X, A{\cdot}B{\cdot}C}\}_n$; $\Psi_{3b_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A{\cdot}B{\cdot}C}\}_n$; $\Psi_{3b_n}^{\text{exclude}} \equiv \emptyset$

Note: Step 3 combining an assembly operation (Step 3a; addition of C) with a disassembly operation (Step 3b; removal of X) is described using two target test tubes; the Step 3a tube prevents completion of the full operation by excluding the final product A·B·C from the ensemble ($L_{\max} = 2$ includes all off-targets up to dimers).

Crosstalk tube

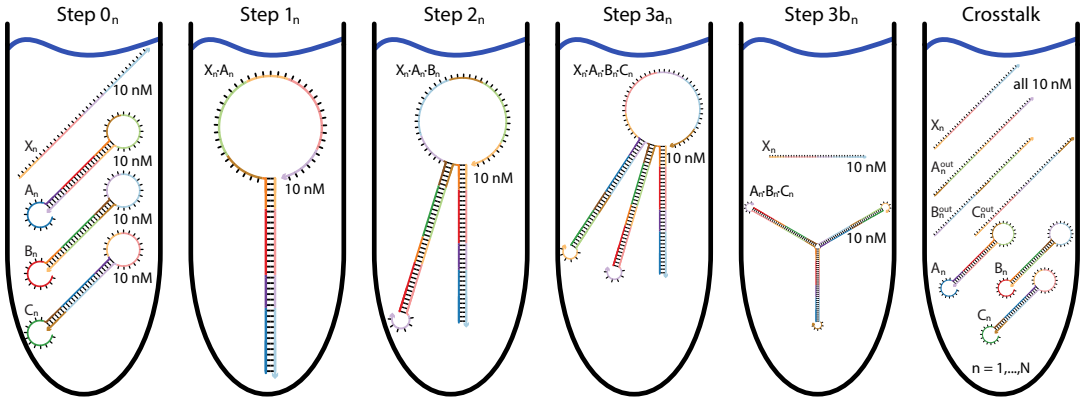
- Crosstalk tube: $\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$; $\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L \le L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

The reactive species and cognate products for system n are:

- $\lambda_n^{\text{simple}} \equiv \{\mathrm{A, B, C}\}_n$
- $\lambda_n^{\text{ss-out}} \equiv \{\mathrm{X, A^{out}, B^{out}, C^{out}}\}_n$
- $\lambda_n^{\text{ss-in}} \equiv \{\mathrm{A^{toe}, B^{toe}, C^{toe}}\}_n$
- $\lambda_n^{\text{reactive}} \equiv \{\mathrm{A, B, C, X, A^{out}, B^{out}, C^{out}}\}_n$
- $\lambda_n^{\text{cognate}} \equiv \{\mathrm{X{\cdot}A, A^{out}{\cdot}B, B^{out}{\cdot}C, C^{out}{\cdot}A, C^{out}{\cdot}B, B^{out}{\cdot}A, A^{out}{\cdot}C}\}_n$

based on the definitions (listed 5′ to 3′ using the sequence domain notation of Figure S3):

- $\mathrm{A} \equiv \mathrm{A^{in}\text{-}A^{out}}$
- $\mathrm{A^{toe}} \equiv \mathrm{a}$
- $\mathrm{A^{in}} \equiv \mathrm{a\text{-}x\text{-}b\text{-}y}$
- $\mathrm{A^{out}} \equiv \mathrm{z^*\text{-}c^*\text{-}y^*\text{-}b^*\text{-}x^*}$
- $\mathrm{B} \equiv \mathrm{B^{in}\text{-}B^{out}}$
- $\mathrm{B^{toe}} \equiv \mathrm{b}$
- $\mathrm{B^{in}} \equiv \mathrm{b\text{-}y\text{-}c\text{-}z}$
- $\mathrm{B^{out}} \equiv \mathrm{x^*\text{-}a^*\text{-}z^*\text{-}c^*\text{-}y^*}$
- $\mathrm{C} \equiv \mathrm{C^{in}\text{-}C^{out}}$
- $\mathrm{C^{toe}} \equiv \mathrm{c}$
- $\mathrm{C^{in}} \equiv \mathrm{c\text{-}z\text{-}a\text{-}x}$
- $\mathrm{C^{out}} \equiv \mathrm{y^*\text{-}b^*\text{-}x^*\text{-}a^*\text{-}z^*}$
- $\mathrm{X} \equiv \mathrm{y^*\text{-}b^*\text{-}x^*\text{-}a^*}$



Tube	On-targets (Ψ_h^{on})	Off-targets (Ψ_h^{off})
Step 0_n	$\{\text{X, A, B, C}\}_n$	$\Psi_{0_n}^{L\leq L_{\max}} - \{\text{X}\cdot\text{A}\}_n$
Step 1_n	$\{\text{X}\cdot\text{A}\}_n$	$\{\text{X, A}\}_n \cup \Psi_{1_n}^{L\leq L_{\max}}$
Step 2_n	$\{\text{X}\cdot\text{A}\cdot\text{B}\}_n$	$\{\text{X}\cdot\text{A, B}\}_n \cup \Psi_{2_n}^{L\leq L_{\max}}$
Step $3a_n$	$\{\text{X}\cdot\text{A}\cdot\text{B}\cdot\text{C}\}_n$	$\{\text{X}\cdot\text{A}\cdot\text{B, C}\}_n \cup \Psi_{3a_n}^{L\leq L_{\max}}$
Step $3b_n$	$\{\text{X, A}\cdot\text{B}\cdot\text{C}\}_n$	$\{\text{X}\cdot\text{A}\cdot\text{B}\cdot\text{C}\}_n \cup \Psi_{3b_n}^{L\leq L_{\max}}$
Crosstalk	$\cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$	$\Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

Figure S8. Target test tubes for self-assembly of a 3-arm junction via CHA (reaction pathway of Figure S3). Top: Target test tube schematics. Bottom: Target test tube details. Each target test tube contains the depicted on-target complexes (each with the depicted target structure and a target concentration of 10 nM) and the off-target complexes listed in the table (each with vanishing target concentration). To simultaneously design N orthogonal systems, the total number of target test tubes is $|\Omega| = 5N + 1$. $L_{\max} = 2$ for all tubes. Design conditions: DNA in 1 M Na^+ at 25 °C.

S2.2.5 Boolean Logic AND using a Cooperative Hybridization Gate

Target test tubes are defined using the specification of Section S2.2.1 with the following definitions. The total number of target test tubes is $|\Omega| = \sum_{n=1,\ldots,N} \{\text{Step 0, Step 1}\}_n$ + Crosstalk = $2N+1$; the target test tubes in the multistate test tube ensemble, Ω, are indexed by $h = 1, \ldots, 2N+1$. $L_{\max} = 2$ for all tubes.

Reactants for system n

- Targets: $\{\text{X, Y}\}_n$
- Cooperative gate: $\{\text{A·B}\}_n$

Elementary step tubes for system n

- Step 0_n: $\Psi_{0_n}^{\text{products}} \equiv \{\text{X, Y, A·B}\}_n$; $\Psi_{0_n}^{\text{reactants}} \equiv \{\text{A·X·B, A·Y·B}\}_n$; $\Psi_{0_n}^{\text{exclude}} \equiv \emptyset$
- Step 1_n: $\Psi_{1_n}^{\text{products}} \equiv \{\text{Y·A·X, B}\}_n$; $\Psi_{1_n}^{\text{reactants}} \equiv \{\text{X, Y, A·B}\}_n$; $\Psi_{1_n}^{\text{exclude}} \equiv \emptyset$

Note: Individual targets do not appreciably bind the cooperative gate. In the Step 0_n tube, the reactants are prevented from generating the product Y·A·X, because this trimer is excluded from the test tube ensemble ($L_{\max} = 2$ includes all off-targets up to dimers).

Crosstalk tube

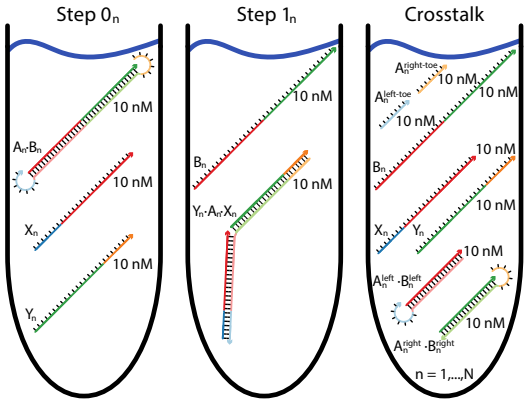
- Crosstalk tube: $\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$; $\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L \le L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

The reactive species and cognate products for system n are:

- $\lambda_n^{\text{simple}} \equiv \{\text{A}^{\text{left}}\text{·B}^{\text{left}}, \text{A}^{\text{right}}\text{·B}^{\text{right}}\}_n$
- $\lambda_n^{\text{ss-out}} \equiv \{\text{X, Y, B}\}_n$
- $\lambda_n^{\text{ss-in}} \equiv \{\text{A}^{\text{left-toe}}, \text{A}^{\text{right-toe}}\}_n$
- $\lambda_n^{\text{reactive}} \equiv \{\text{A}^{\text{left}}\text{·B}^{\text{left}}, \text{A}^{\text{right}}\text{·B}^{\text{right}}, \text{X, Y, B, A}^{\text{left-toe}}, \text{A}^{\text{right-toe}}\}_n$
- $\lambda_n^{\text{cognate}} \equiv \{\text{X·A}^{\text{left}}, \text{Y·A}^{\text{right}}, \text{X·A}^{\text{left-toe}}, \text{Y·A}^{\text{right-toe}}, \text{A}^{\text{left}}\text{·B, A}^{\text{right}}\text{·B}\}_n$

based on the definitions (listed 5′ to 3′ using the sequence domain notation of Figure S4):

- $\text{A} \equiv \text{A}^{\text{right}}\text{-A}^{\text{left}}$
- $\text{A}^{\text{left-toe}} \equiv \text{a*}$
- $\text{A}^{\text{left}} \equiv \text{b*-a*}$
- $\text{A}^{\text{right-toe}} \equiv \text{d*}$
- $\text{A}^{\text{right}} \equiv \text{d*-c*}$
- $\text{B} \equiv \text{B}^{\text{left}}\text{-B}^{\text{right}}$
- $\text{B}^{\text{left}} \equiv \text{b}$
- $\text{B}^{\text{right}} \equiv \text{c}$
- $\text{X} \equiv \text{a-b}$
- $\text{Y} \equiv \text{c-d}$



Tube	On-targets (Ψ_h^{on})	Off-targets (Ψ_h^{off})
Step 0_n	$\{\text{X, Y, A·B}\}_n$	$\{\text{A·X·B, A·Y·B}\}_n \cup \Psi_{0_n}^{L \leq L_{\max}}$
Step 1_n	$\{\text{Y·A·X, B}\}_n$	$\{\text{X, Y, A·B}\}_n \cup \Psi_{1_n}^{L \leq L_{\max}}$
Crosstalk	$\cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$	$\Psi_{\text{global}}^{L \leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

Figure S9. Target test tubes for Boolean logic AND using a cooperative hybridization gate (reaction pathway of Figure S4). Top: Target test tube schematics. Bottom: Target test tube details. Each target test tube contains the depicted on-target complexes (each with the depicted target structure and a target concentration of 10 nM) and the off-target complexes listed in the table (each with vanishing target concentration). To simultaneously design N orthogonal systems, the total number of target test tubes is $|\Omega| = 2N + 1$. $L_{\max} = 2$ for all tubes. Design conditions: DNA in 1 M Na^+ at 25 °C.

S2.2.6 Conditional Dicer Substrate Formation via Shape and Sequence Transduction with scRNAs

Target test tubes are defined using the specification of Section S2.2.1 with the following definitions. The total number of target test tubes is $|\Omega| = \sum_{n=1,\ldots,N} \{\text{Step 0, Step 1, Step 2}\}_n$ + Crosstalk = $3N + 1$; the target test tubes in the multistate test tube ensemble, Ω, are indexed by $h = 1, \ldots, 3N + 1$. $L_{\max} = 2$ for all tubes.

Reactants for system n

- Target: X_n
- scRNAs: $\{\mathrm{A{\cdot}B},\ \mathrm{C}\}_n$

Elementary step tubes for system n

- Step 0_n: $\Psi_{0_n}^{\text{products}} \equiv \{\mathrm{X},\ \mathrm{A{\cdot}B},\ \mathrm{C}\}_n$; $\Psi_{0_n}^{\text{reactants}} \equiv \{\mathrm{A},\ \mathrm{B{\cdot}C}\}_n$; $\Psi_{0_n}^{\text{exclude}} \equiv \{\mathrm{X{\cdot}A}\}$
- Step 1_n: $\Psi_{1_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A},\ \mathrm{B}\}_n$; $\Psi_{1_n}^{\text{reactants}} \equiv \{\mathrm{X},\ \mathrm{A{\cdot}B}\}_n$; $\Psi_{1_n}^{\text{exclude}} \equiv \emptyset$
- Step 2_n: $\Psi_{2_n}^{\text{products}} \equiv \{\mathrm{B{\cdot}C}\}_n$; $\Psi_{2_n}^{\text{reactants}} \equiv \{\mathrm{B},\ \mathrm{C}\}_n$; $\Psi_{2_n}^{\text{exclude}} \equiv \emptyset$

Crosstalk tube

- Crosstalk tube: $\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$; $\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L \le L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

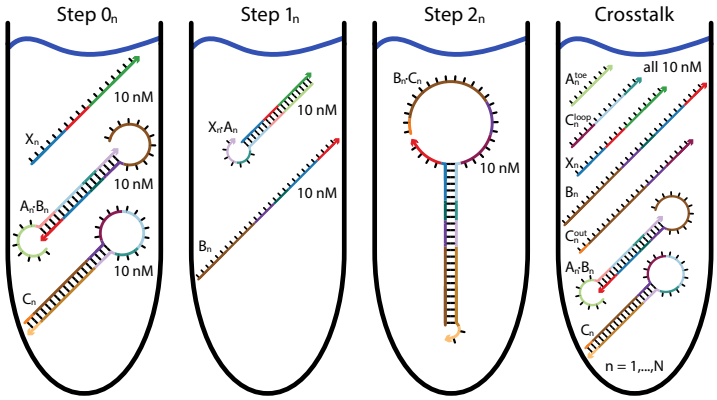
The reactive species and cognate products for system n are:

- $\lambda_n^{\text{simple}} = \{\mathrm{A{\cdot}B},\ \mathrm{C}\}_n$
- $\lambda_n^{\text{ss-out}} = \{\mathrm{X},\ \mathrm{B},\ \mathrm{C^{out}}\}_n$
- $\lambda_n^{\text{ss-in}} = \{\mathrm{A^{toe}},\ \mathrm{C^{loop}}\}_n$
- $\lambda_n^{\text{reactive}} = \{\mathrm{A{\cdot}B},\ \mathrm{C},\ \mathrm{X},\ \mathrm{B},\ \mathrm{C^{out}},\ \mathrm{A^{toe}},\ \mathrm{C^{loop}}\}_n$
- $\lambda_n^{\text{cognate}} = \{\mathrm{X{\cdot}A},\ \mathrm{B{\cdot}C},\ \mathrm{X{\cdot}A^{toe}},\ \mathrm{B{\cdot}C^{loop}}\}_n$

based on the definitions (listed 5′ to 3′ using the sequence domain notation of Figure S5):

- $\mathrm{A} \equiv$ c*-b*-a*-z*-y*
- $\mathrm{A^{toe}} \equiv$ c*
- $\mathrm{B} \equiv$ x-y-z-a-b
- $\mathrm{C} \equiv \mathrm{C^{out}}$-$\mathrm{C^{in}}$
- $\mathrm{C^{loop}} \equiv$ s-a*-z*
- $\mathrm{C^{in}} \equiv$ a*-z*-y*-x*-w*
- $\mathrm{C^{out}} \equiv$ w-x-y-s
- $\mathrm{X} \equiv$ a-b-c

Note: $\mathrm{C}_n^{\text{loop}}$ includes portions of both C_n^{in} and $\mathrm{C}_n^{\text{out}}$. Including $\mathrm{C}_n^{\text{loop}}$ in $\lambda_n^{\text{reactive}}$ is not redundant with inclusion of C_n because pairing to the loop would cause a pseudoknot and hence will not be checked by the ensemble except if the interaction opens the hairpin. We want to be able to check nucleation with the loop even when the hairpin remains closed, so we include $\mathrm{C}_n^{\text{loop}}$ in $\lambda_n^{\text{reactive}}$.



Tube	On-targets (Ψ_h^{on})	Off-targets (Ψ_h^{off})
Step 0_n	$\{\text{X, A·B, C}\}_n$	$\{\text{A, B·C}\}_n \cup \Psi_{0_n}^{L\leq L_{\max}} - \{\text{X·A}\}$
Step 1_n	$\{\text{X·A, B}\}_n$	$\{\text{X, A·B}\}_n \cup \Psi_{1_n}^{L\leq L_{\max}}$
Step 2_n	$\{\text{B·C}\}_n$	$\{\text{B, C}\}_n \cup \Psi_{2_n}^{L\leq L_{\max}}$
Crosstalk	$\cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$	$\Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

Figure S10. Target test tubes for conditional Dicer substrate formation via shape and sequence transduction with scRNAs (reaction pathway of Figure S5). Top: Target test tube schematics. Bottom: Target test tube details. Each target test tube contains the depicted on-target complexes (each with the depicted target structure and a target concentration of 10 nM) and the off-target complexes listed in the table (each with vanishing target concentration). To simultaneously design N orthogonal systems, the total number of target test tubes is $|\Omega| = 3N + 1$. $L_{\max} = 2$ for all tubes. Design conditions: RNA in 1 M Na^+ at 37 °C.

S2.3 Algorithm Performance

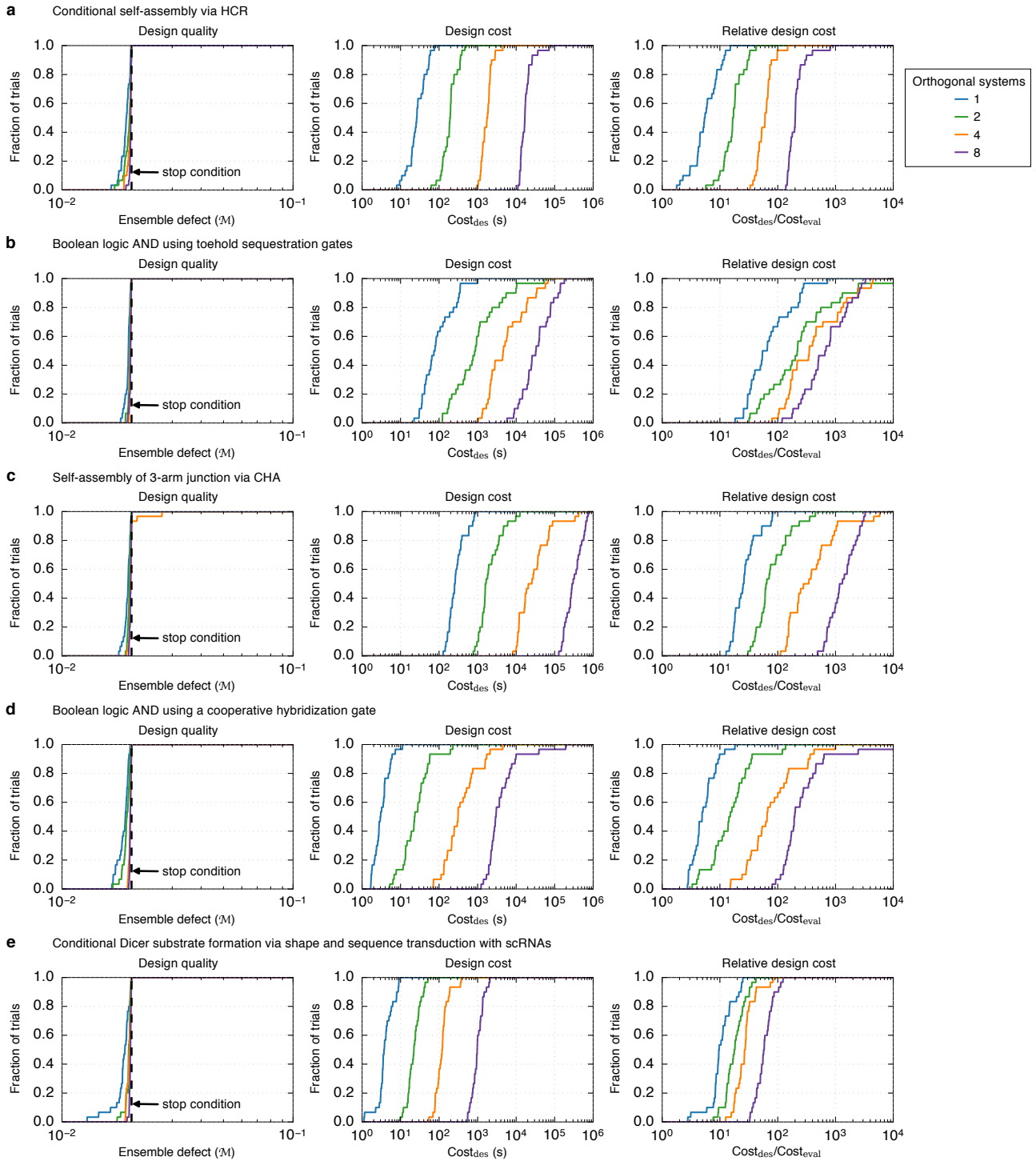


Figure S11. Algorithm performance for design of 1, 2, 4, or 8 orthogonal systems based on the target test tubes of Section S2.2. Left: Design quality. The stop condition is depicted as a dashed black line. Middle: Design cost. Right: Cost of sequence design relative to a single evaluation of the objective function. (a) Conditional self-assembly via HCR. (b) Boolean logic AND using toehold sequestration gates. (c) Self-assembly of 3-arm junction via CHA. (d) Boolean logic AND using a cooperative hybridization gate. (e) Conditional Dicer substrate formation via shape and sequence transduction with scRNAs.

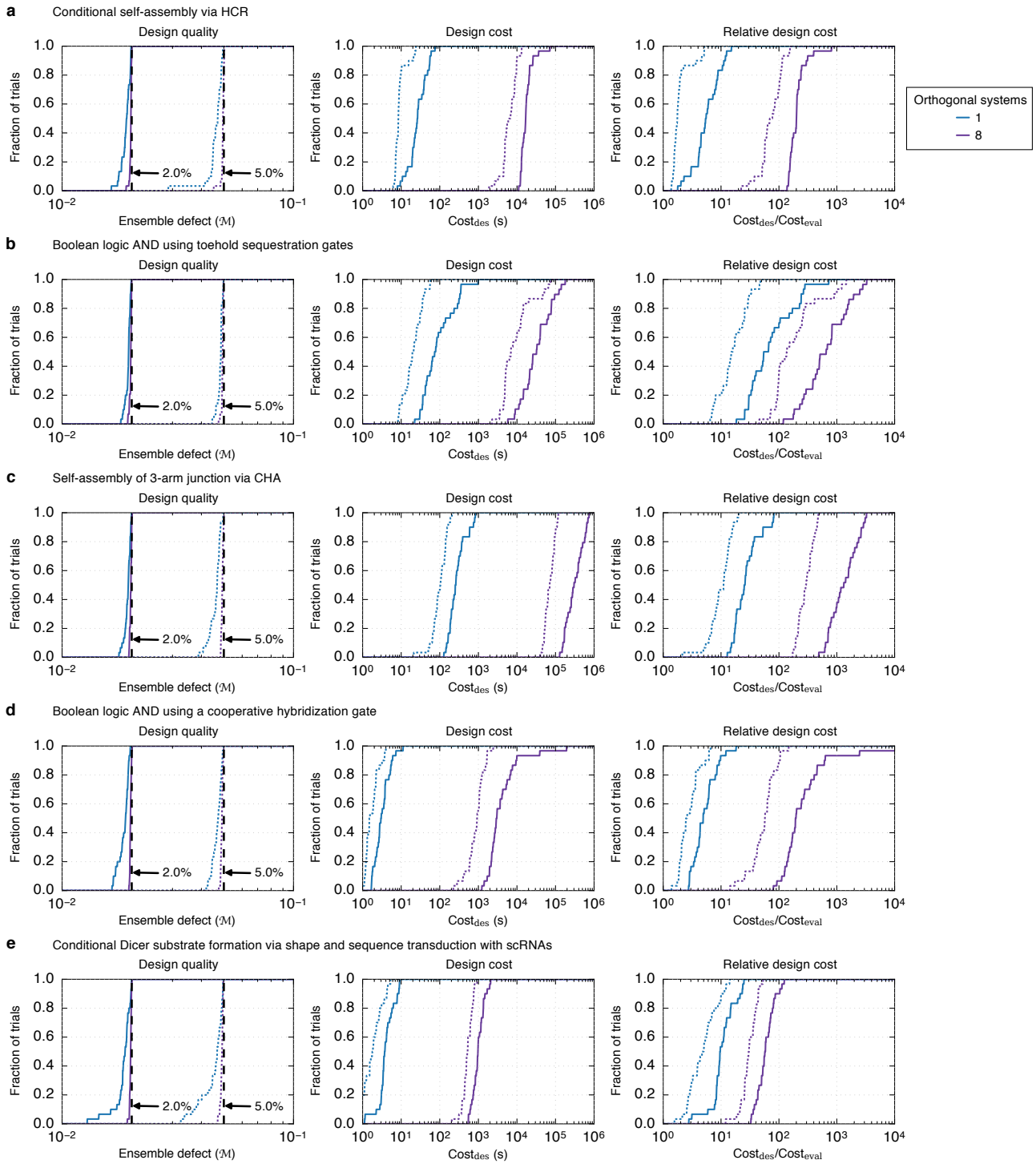


Figure S12. Reduced design cost and quality using $f_{\text{stop}} = 0.05$ (dotted lines) instead of $f_{\text{stop}} = 0.02$ (solid lines) for design of 1 or 8 orthogonal systems based on the target test tubes of Section S2.2. Left: Design quality. The stop conditions are depicted as dashed black lines. Middle: Design cost. Right: Cost of sequence design relative to a single evaluation of the objective function. (a) Conditional self-assembly via HCR. (b) Boolean logic AND using toehold sequestration gates. (c) Self-assembly of 3-arm junction via CHA. (d) Boolean logic AND using a cooperative hybridization gate. (e) Conditional Dicer substrate formation via shape and sequence transduction with scRNAs.

S2.4 Residual Defects

For each case study, the residual defect plots that follow display for each target test tube $h \in \Omega$:

- The structural defect and concentration defect contributions of each on-target complex to the test tube ensemble defect, $\mathcal{M}_h$. The fact that bars are not depicted for off-targets does not mean that the defects associated with off-targets are neglected. By conservation of mass, nonzero off-target concentrations imply deficiencies in on-target concentrations, and these on-target concentration defects are depicted in the bar graphs and incorporated in $\mathcal{M}_h$ via (1) and (5).
- The total structural defect and total concentration defect contributions of the tube to the multistate test tube ensemble defect, $\mathcal{M}$.

S2.4.1 Conditional Self-Assembly via HCR

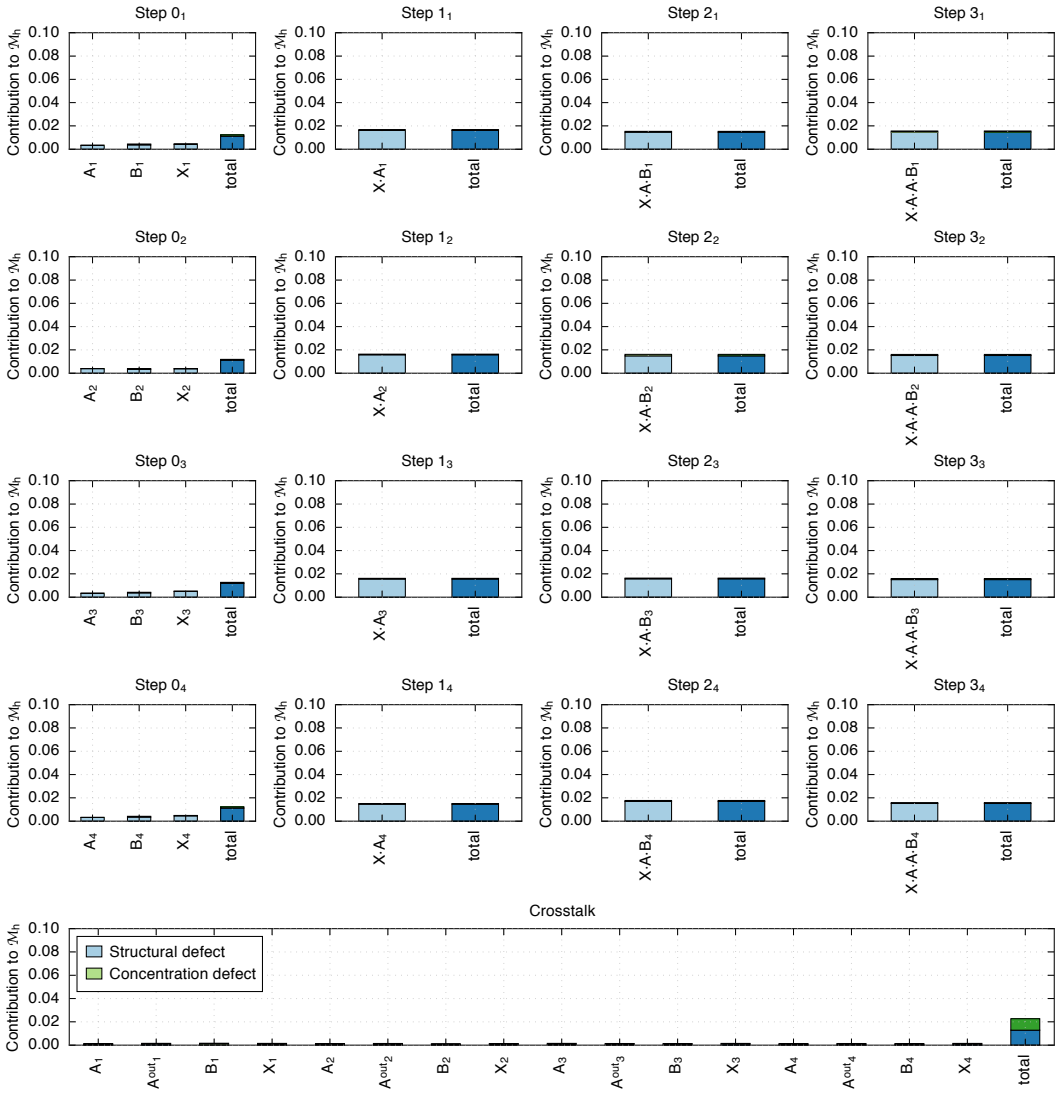


Figure S13. Residual defects for conditional self-assembly via HCR ($N = 4$ orthogonal systems for target test tubes of Figure S6). Each panel corresponds to a different tube $h \in \{1, \ldots, 4N + 1\}$. For each tube h, the structural defect and concentration defect contributions to the tube ensemble defect, $\mathcal{M}_h$, are depicted for each complex (pale shaded bars). The total structural defect and total concentration defect contributions to the multistate test tube ensemble defect, $\mathcal{M}$, are also depicted for each tube (dark shaded bars). Each bar represents the mean over 30 independent design trials with stop condition $\mathcal{M} \leq 0.02$. All nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases.

S2.4.2 Boolean Logic AND using Toehold Sequestration Gates

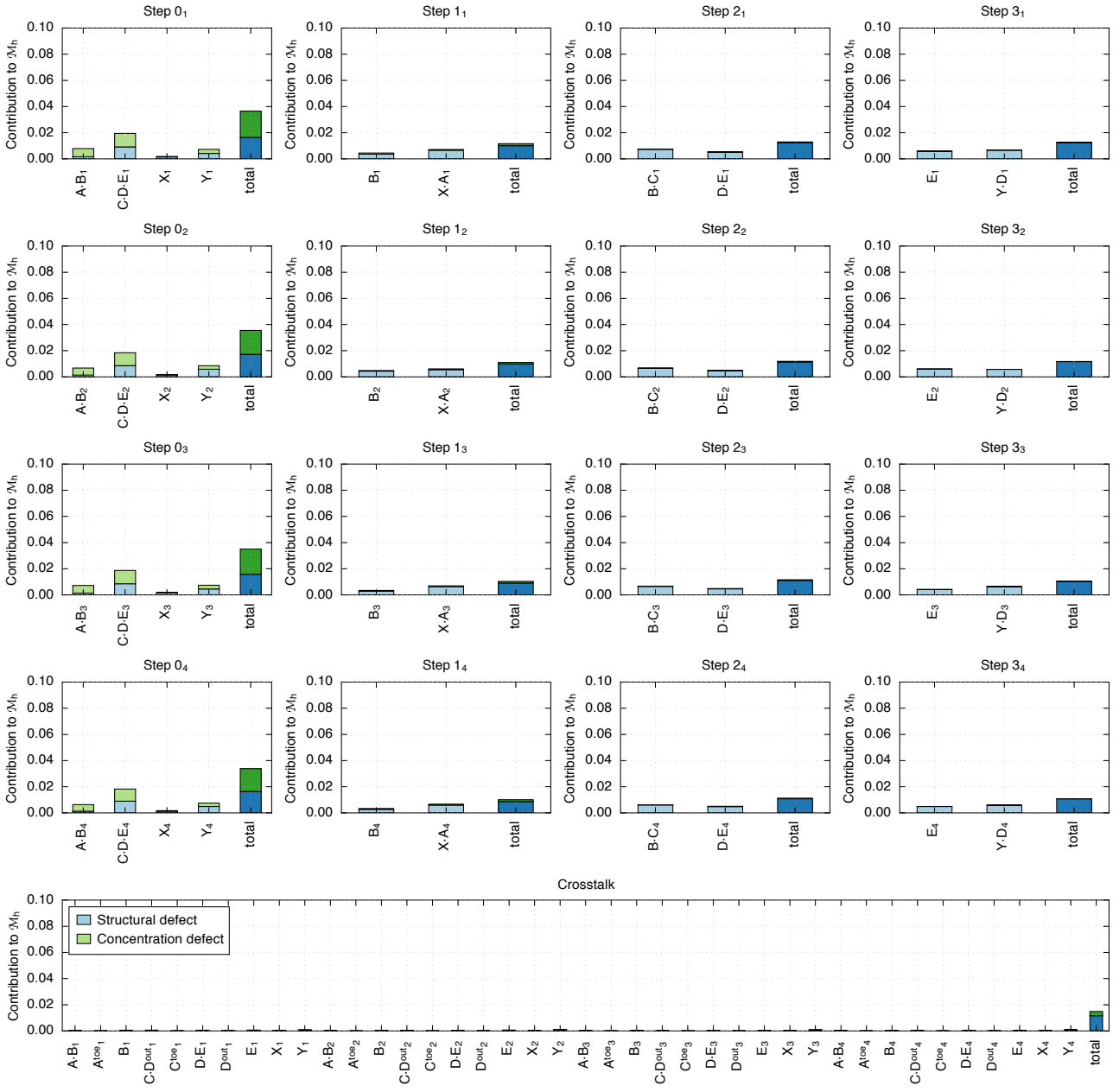


Figure S14. Residual defects for Boolean logic AND using toehold sequestration gates ($N = 4$ orthogonal systems for target test tubes of Figure S7). Each panel corresponds to a different tube $h \in \{1, \ldots, 4N + 1\}$. For each tube h, the structural defect and concentration defect contributions to the tube ensemble defect, $\mathcal{M}_h$, are depicted for each complex (pale shaded bars). The total structural defect and total concentration defect contributions to the multistate test tube ensemble defect, $\mathcal{M}$, are also depicted for each tube (dark shaded bars). Each bar represents the mean over 30 independent design trials with stop condition $\mathcal{M} \leq 0.02$. All nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases.

S2.4.3 Self-Assembly of a 3-Arm Junction via CHA

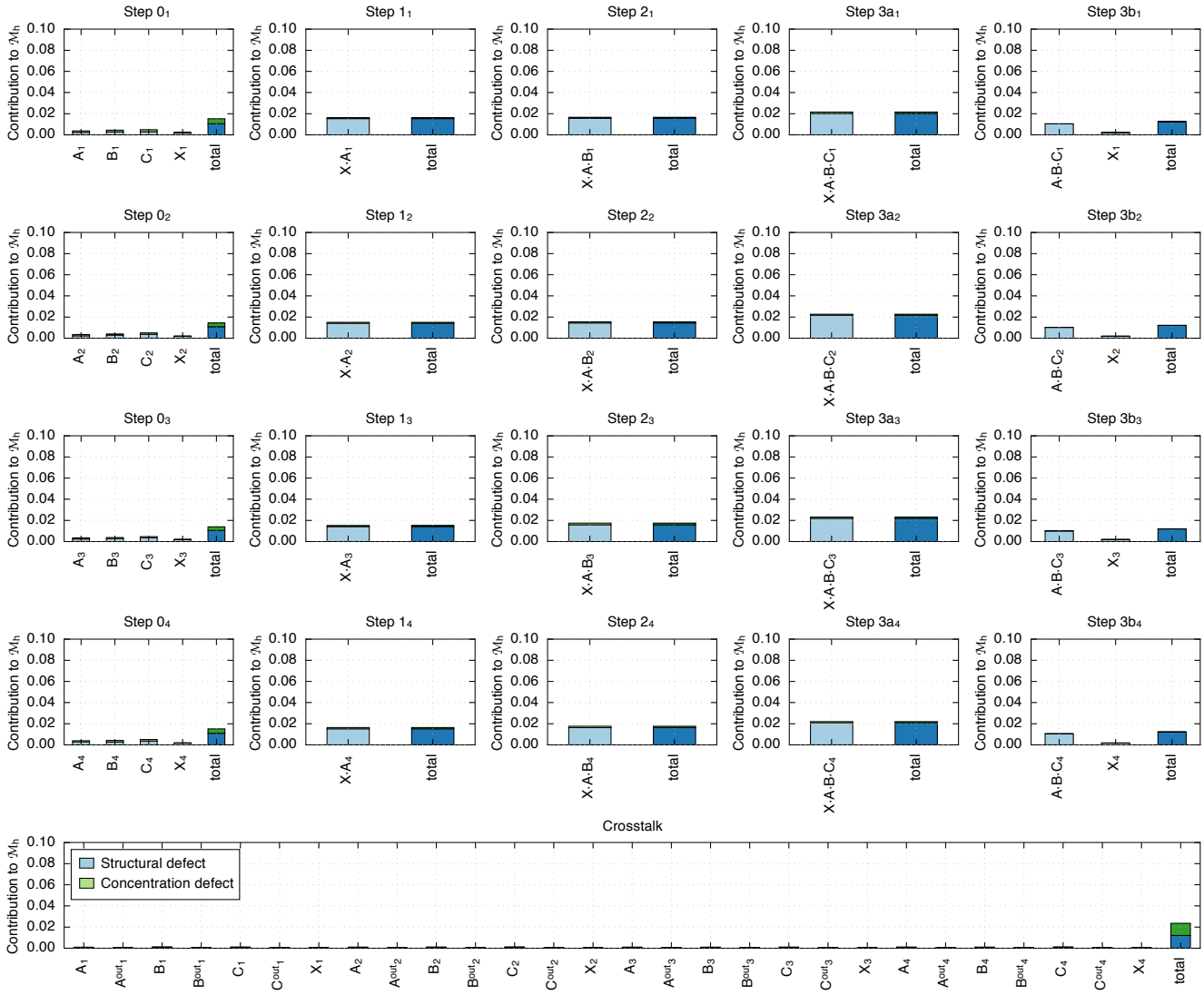


Figure S15. Residual defects for self-assembly of a 3-arm junction via CHA ($N = 4$ orthogonal systems for target test tubes of Figure S8). Each panel corresponds to a different tube $h \in \{1, \ldots, 5N + 1\}$. For each tube h, the structural defect and concentration defect contributions to the tube ensemble defect, $\mathcal{M}_h$, are depicted for each complex (pale shaded bars). The total structural defect and total concentration defect contributions to the multistate test tube ensemble defect, $\mathcal{M}$, are also depicted for each tube (dark shaded bars). Each bar represents the mean over 30 independent design trials with stop condition $\mathcal{M} \leq 0.02$. All nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases.

S2.4.4 Boolean Logic AND using a Cooperative Hybridization Gate

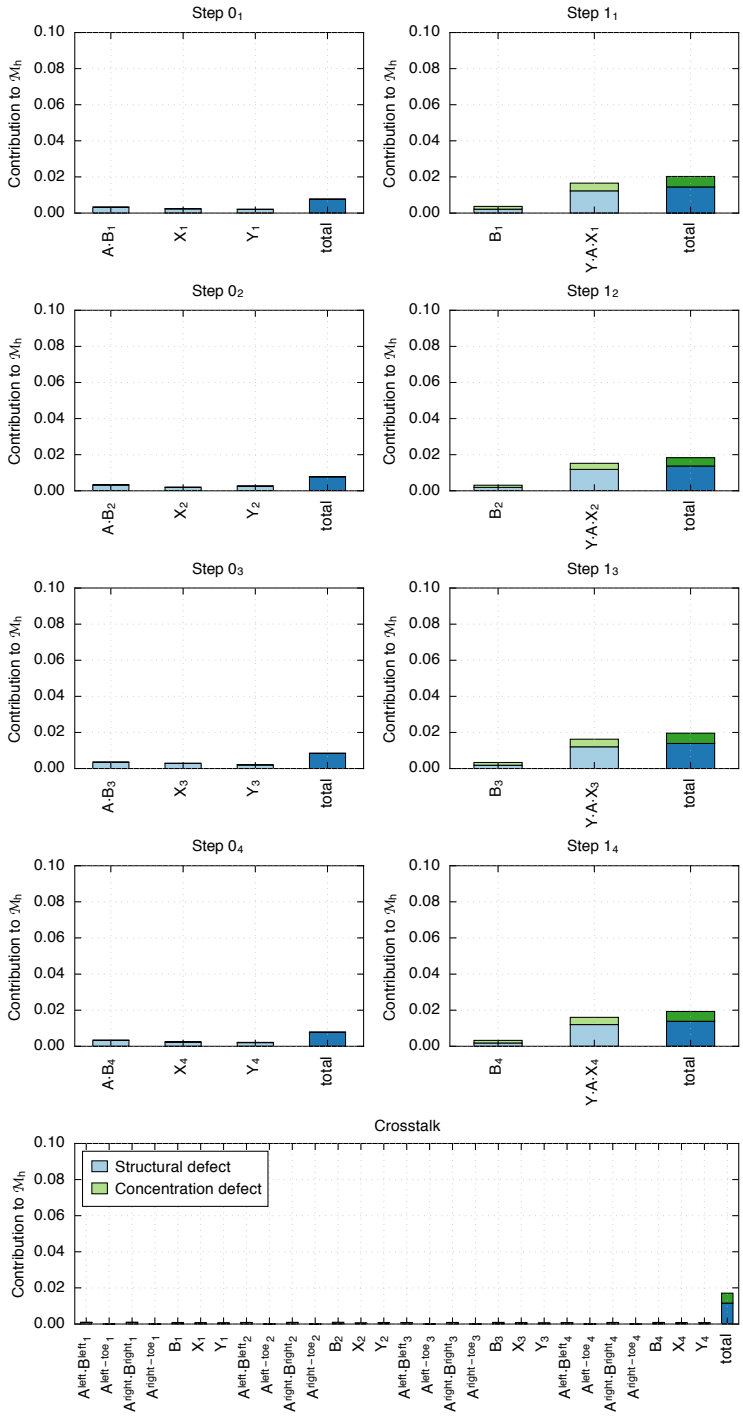


Figure S16. Residual defects for Boolean logic AND using a cooperative hybridization gate ($N = 4$ orthogonal systems for target test tubes of Figure S9). Each panel corresponds to a different tube $h \in \{1, \ldots, 2N + 1\}$. For each tube h, the structural defect and concentration defect contributions to the tube ensemble defect, $\mathcal{M}_h$, are depicted for each complex (pale shaded bars). The total structural defect and total concentration defect contributions to the multistate test tube ensemble defect, $\mathcal{M}$, are also depicted for each tube (dark shaded bars). Each bar represents the mean over 30 independent design trials with stop condition $\mathcal{M} \leq 0.02$. All nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases.

S2.4.5 Conditional Dicer Substrate Formation via Shape and Sequence Transduction with scRNAs

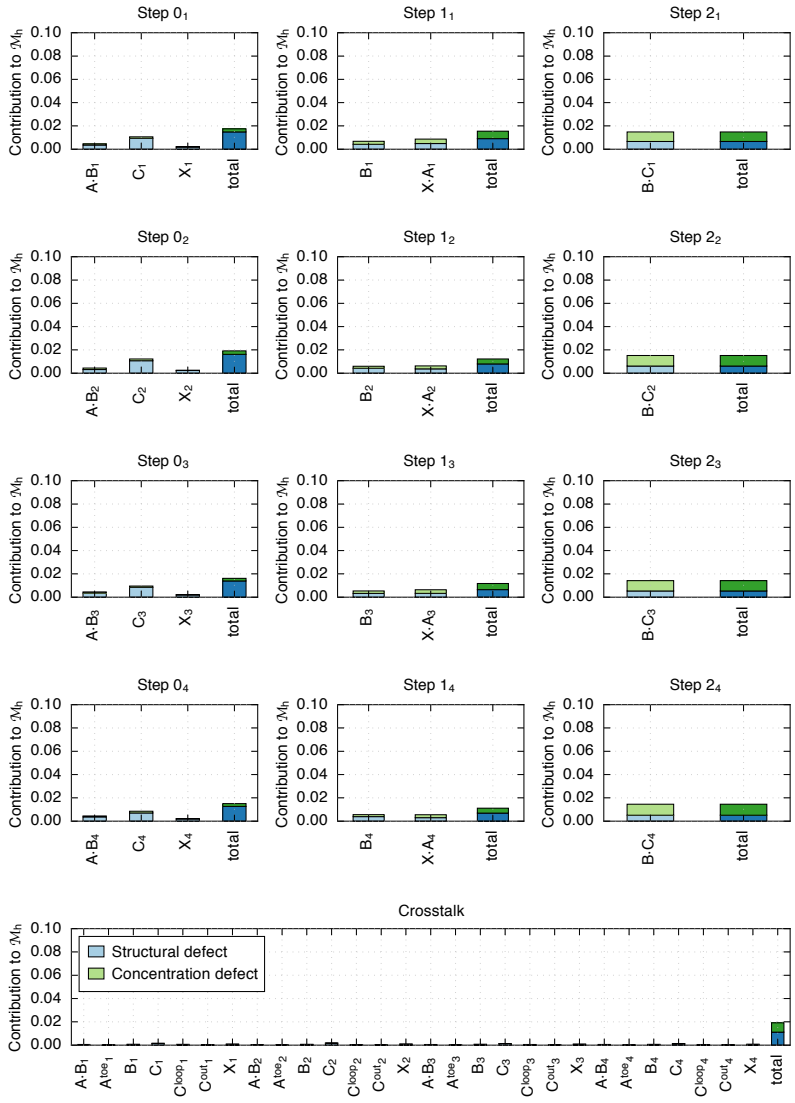


Figure S17. Residual defects for conditional Dicer substrate formation via shape and sequence transduction with scRNAs ($N = 4$ orthogonal systems for target test tubes of Figure S10). Each panel corresponds to a different tube $h \in \{1, \ldots, 3N+1\}$. For each tube h, the structural defect and concentration defect contributions to the tube ensemble defect, $\mathcal{M}_h$, are depicted for each complex (pale shaded bars). The total structural defect and total concentration defect contributions to the multistate test tube ensemble defect, $\mathcal{M}$, are also depicted for each tube (dark shaded bars). Each bar represents the mean over 30 independent design trials with stop condition $\mathcal{M} \leq 0.02$. All nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases.

S2.5 Importance of Negative Design in Reducing Crosstalk

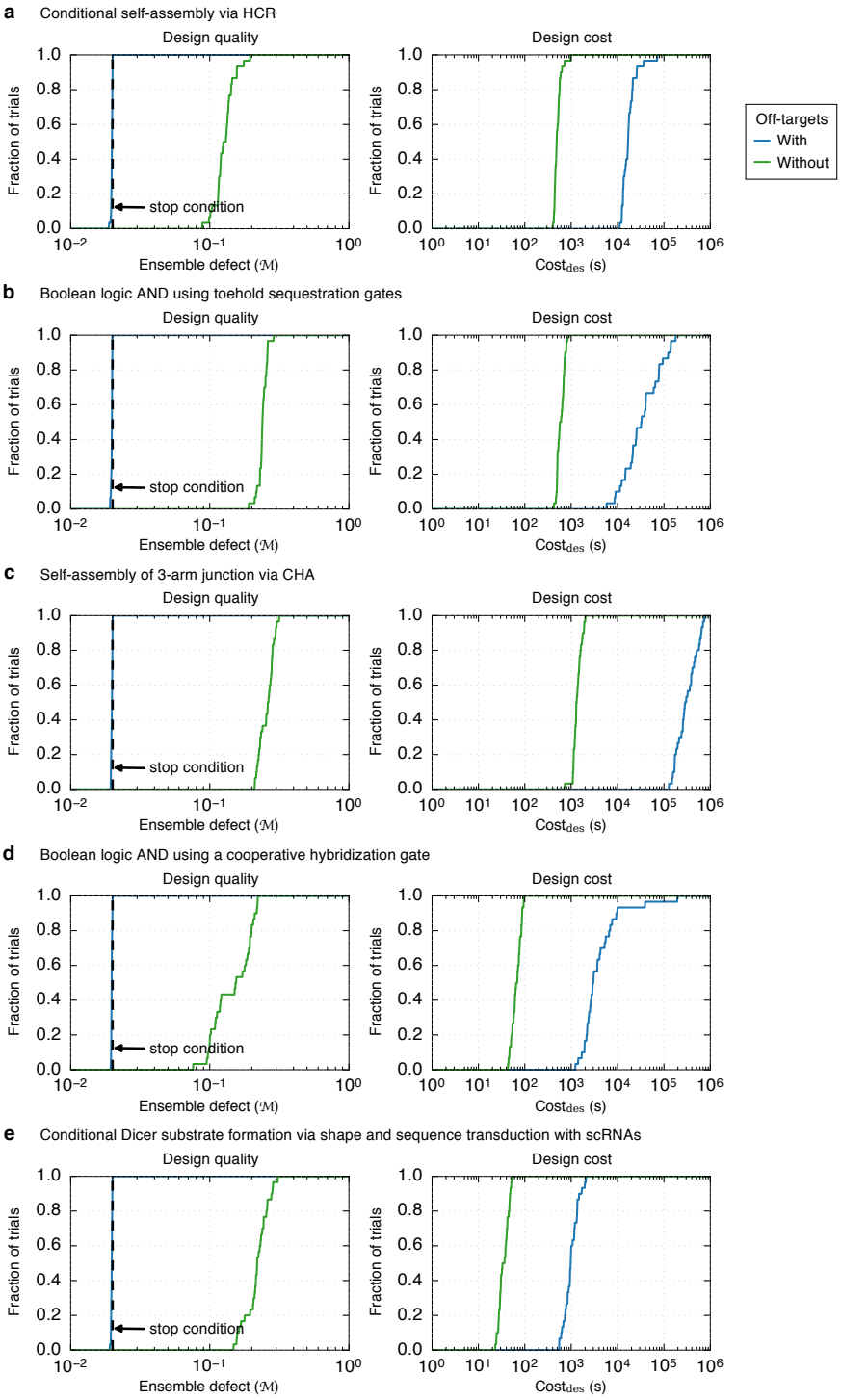


Figure S18. Importance of negative design in reducing crosstalk ($N = 8$ orthogonal systems). Comparison of designs performed with or without off-targets in the design ensemble. Left: Design quality evaluated by calculating the multistate test tube ensemble defect ($\mathcal{M}$) over the ensemble containing off-targets. The stop condition is depicted as a dashed black line. Right: Design cost. (a) Conditional self-assembly via HCR. (b) Boolean logic AND using toehold sequestration gates. (c) Self-assembly of 3-arm junction via CHA. (d) Boolean logic AND using a cooperative hybridization gate. (e) Conditional Dicer substrate formation via shape and sequence transduction with scRNAs.

S2.6 Effect of Sequence Constraints

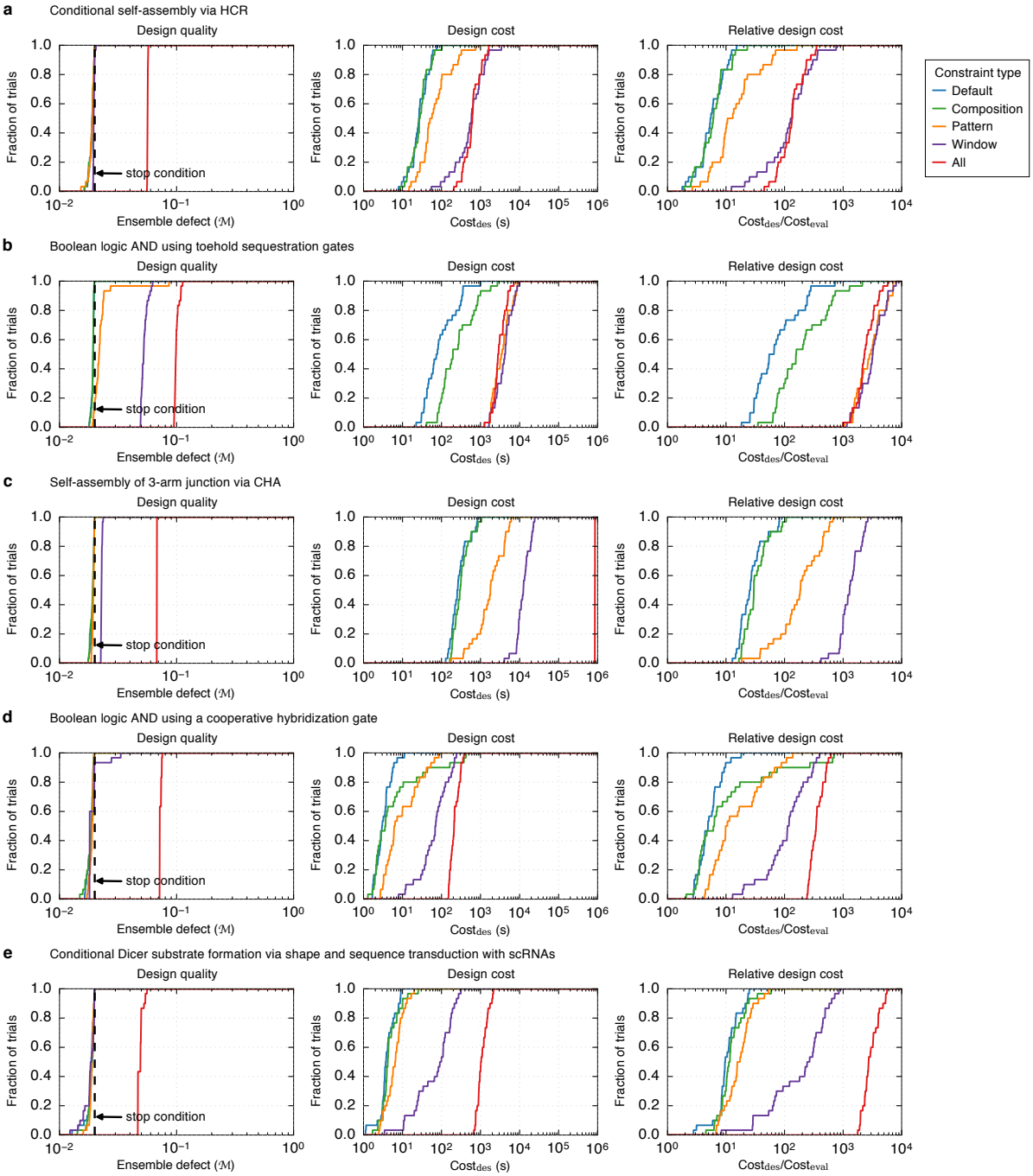


Figure S19. Algorithm performance including explicit sequence constraints ($N = 1$ system). Default: implicit sequence constraints inherent to the reaction pathway (these constraints are also present in the other cases that follow). Composition constraint: fraction of S $\in [0.45, 0.55]$. Pattern constraint: prevent {AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY}. Window constraints: targets X and Y constrained to be subsequences of two different mRNAs (i.e., biological sequence constraints; see Section S2.6.1). All: all of the above constraints. Left: Design quality. The stop condition is depicted as a dashed black line. Middle: Design cost. Right: Cost of sequence design relative to a single evaluation of the objective function. (a) Conditional self-assembly via HCR. (b) Boolean logic AND using toehold sequestration gates. (c) Self-assembly of 3-arm junction via CHA. (d) Boolean logic AND using a cooperative hybridization gate. (e) Conditional Dicer substrate formation via shape and sequence transduction with scRNAs. Note: for panel (c) with "All constraints", only 1 out of 30 independent design trials terminates within the depicted time frame.

S2.6.1 mRNA Sequences used for Window Constraints

The window constraints used for the studies of Figures 5 and S19 constrain targets X and Y to be subsequences of zebrafish mRNA sequences *tpm3* (2175 nt) and *desma* (1798 nt), respectively. Zebrafish mRNA sequences were obtained from the National Center for Biotechnology Information (NCBI).[61] Sequences are listed below $5'$ to $3'$. Table S3 lists the sequence domains constrained by window constraints for each reaction pathway, the number of candidate windows within each mRNA, and the number of these candidate windows that also satisfy the composition constraints, the pattern prevention constraints, or both the composition and pattern prevention constraints.

mRNA: **tropomyosin 3** (*tpm3*)
Sequence: gAACACUAUUAgCUAUUUgUAgUACUCUAAAgAggACUgCAgAACgCAUCgCAgUAgUggUgAAAAgCCgUgCgUgCgCgUgAAACAUCUgAUCC
UCACgUUACUUCCCACUCgCUCUgCgUUUgACUUgUUggCggggCgUUggUgCCUUggACUUUUUUUUCCUCCUUCUCUUCUUCgCggCUCggUCCACUACgCUgC
UCgAgAggAAUCUgCUUUAUUCgACCACACUACUCCUAAAgUAACACAUUAAAAUggCCggAUCAAACAgCAUCgAUgCAgUUAAgAgAAAAAUCAAAgUUUUAC
AACAgCAAgCAgAUgAggCAgAAgAAAgAgCCgAgAUUUUgCAgAgACAggUCgAggAggAgAAgCgUgCCAgggAgCAggCUgAggCAgAggUggCUUCUCUgA
ACAggCgUAUCCAgCUggUUgAggAggAgUUggAUCgUgCUCAggAgAgACUggCCACAgCCCUgCAAAAgCUggAggAAgCCgAgAAggCCgCAgAUgAgAgCg
AgAgAgggAUgAAggUgAUUgAgAACAgggCUCUgAAggAUgAggAgAAgAUggAgCUgCAggAgAUCCAgCUUAAggAggCCAAgCACAUUgCUgAggAggCUg
ACCgCAAAUAUgAAgAggUggCUCgUAAgCUggUgAUCgUUgAgggAgAgUUggAgCgUACAgAggAgAgAgCAgAgCUUgCAgAgAgCCAUgUCAAgCAgAUgg
AggAggAgCUgAgAgCUCUUgACCAgACACUgAAgACUCUUCAggCCUCAgAggAgAAgUAUUCCCAgAAggAggACAAgUAUgAggAAgAAAUCAAgAUCCUCA
CUgAUAAgCUgAAggAggCUgAgACCCgUgCAgAgUUUgCUgAgAggUCUgUggCCAAACUggAgAAAACCAUUgAUgAUUUggAAgAgAAACUgAgAgAUgCUA
AAgAggAgAACAUCAAgAUCCAUgCUACUUUggACCAgACCCUgAgCgAgCUCAAUAgUUUCUAAAgAAgACCUggAgCAgAAAAAAggCCUUUUCUUCCCUUCU
UgACUCCCUCAUCUCAUUUUggUUUCUUUgUCUCUgCACAUCUgAUUCUCCCCCUUUUUUUUUCUUCUCUUCUUCUgCUggAggAUAAgCUCACCAAgCCAACCA
gCAAAAAUgUggUgCCUCUCAAUUUUUCCAAACUACUAUUCCAAgUgAUUUgAgAAAUgAUCUACUACgAUACUCCUCAAgAgUCAAAUgUUgACCUCggggAgC
CUUUUUUggUAUUgCUCCAUgAUCAgAgCUUUACgAgCUAgUgUUUUUUCUgCAUAUCAgCCCAAACUCUCAAUgAUAAUUUUACUggAggCUgAUUUUUgUAAA
AUUUUgUgCCAUAAAAAgCCUUgUUggCUUgUCUCUUgCUUggCUUUAgAUCAUUCUCAAgCCAUUUUUUUCCUgCUgUUgCUCUgACACAggUUgUUUUUgCUgg
UCUUgUUggUgCCUgAUCCACUgCUAUCCUUUUCACACCUCUUUUUUUUUUUUCUUCAUCCUgCACAAgUUUCUgCUgCCUgUUAgUCggCAUCACCggUUUUgg
gACCAAAACCACAUCAUgUggUCUgUAACAgUAUgCACAACCAUgCCgUgAggACCAAAUUUgUUUUAUUAUUgUUAUUAUUAUUAAAAgCCUUUgCUUCCAUUC
ggAgUUUgUUUUUUUgAgUAAUAUAUgUAUUCAUUgUUUgggUCgAAUCCCCUUgCUUUUUUAACACAAAUgUUUUgCAAACCACUAUUUgAAAUggUgCACUgU
UAUgggCUUAUggUgAgCAgAUgAggCCAAgUCAUggUUUCUUCAUUAUAAUUUUCUUUUCAUUUgCUUUAAAgAgCCAUAUUCUACCCAgggAAgAAAggUUgA
AgUUgUUUUgUUUUUUUACCgUgAgUUCAAAgCAgUggCACUgCCAgAUUUAAAAggUUCAAAAgCCgUgCAgAUCUAAAAUAUgUAUUAUgAACACAgUAAUgg
gAgCgAAUUgUAACACUUAAUAgUAUACAAAUUUAAgAAACAggggUgAACACAUAgUUUUAACUggAAAAAgCCCACAAUgAUgUgUAAUCACUUUgUUACUgU
CUgUAUCUUgUgUAAUgAUACCUAAAUUCUUUUUUUAAAUAAAAACCAUgAUUUUUACUgUCACUgAAAAAAAAAAAAAAAAAAAA

mRNA: **desmin a** (*desma*)
Sequence: CAUUUACACAgCgUACAAACCCAACAggCCCAgUCAUgAgCACgAAAUAUUCAgCCUCCgCCgAgUCggCgUCCUCUUACCgCCgCACCUUUggC
UCAggUUUgggCUCCUCUAUUUUCgCCggCCACggUUCCUCAggUUCCUCUggCUCCUCAAgACUgACCUCCAgAgUUUACgAggUgACCAAgAgCUCCgCUUCU
CCCCAUUUUUCCAgCCACCgUgCgUCCggCUCUUUCggAggUggCUCggUggUCCgUUCCUACgCUggCCUUggUgAgAAgCUggAUUUCAAUCUggCUgAUgCC
AUAAACCAggACUUCCUCAACACgCgUACUAAUgAgAAggCCgAgCUCCAgCACCUCAAUgACCgCUUCgCCAgCUACAUCgAgAAggUgCgCUUCCUCgAgCAg
CAgAACUCUgCCCUgACggUggAgAUUgAgCgUCUgCggggUCgCgAgCCCACCCgUAUUgCAgAgCUgUACgAggAggAgAUgAgAgAgCUgCgCggACAggUg
gAggCACUgACCAAUCAgAgAUCCCgUgUggAgAUCgAgAgggACAACCUAgUCgAUgACCUACAgAAACUAAAgCUCAgACUUCAAgAggAgAUCCACCAgAAA
gAggAAgCUgAAAACAACCUUUCUgCUUUCAgAgCUgAUgUCgAUgCUgCCACUCUggCCAggCUggACCUggAAAgACgUAUCgAgggUCUUCACgAAgAgAUU
gCAUUCCUCAggAAgAUUCAUgAggAggAgAUCCgUgAgCUgCAgAACCAgAUgCAggAgAgUCAggUgCAgAUCCAAAUggACAUgUCCAAACCAgACCUgACU
gCggCCCUCAgAgACAUUCgCCUgCAgUACgAggCUAUCgCUgCCAAgAAUAUCAgCgAggCCgAggACUggUAUAAgUCUAAggUUUCAgAUUUgAACCAggCA
gUgAACAAgAAUAACgAggCUCUCAgAgAAgCCAAgCAggAgACCAUgCAgUUCCgUCACCAgCUCCAgUCCUACACCUgCgAgAUUgACUCUCUCAAgggCACC
AAUgAgUCUCUgAggAggCAAAUgAgUgAgAUggAggAgCggCUgggACgUgAggCCggUggUUAUCAggACACUAUCgCCCgUCUCgAggCUgAgAUCgCAAAA
AUgAAAgACgAgAUggCCCgCCACCUCCgCgAgUACCAggAUCUgCUgAAUgUgAAgAUggCUCUggAUgUggAgAUCgCCACCUACAggAAgCUUUUggAAggA
gAggAgAgCAggAUCUCgCUgCCCgUgCAgUCCUUUUCAUCCCUgAgUUUCAgAgAgAgCAgUCCAgAgCAgCACCACCACCAgCAgCAgCAACCACAACgCUCA
UCUgAAgUCCACUCCAAgAAAACAgUCCUgAUCAAgACCAUCgAgACCCgCgAUggCgAggUCgUCAgCgAgUCCACACAgCACCAgCAggACgUCAUgUAAAgC
UUgAgAAACAgAUCgAgUUUCACAgAAUgCCUUgCAUUUUCACUgAUggCCUCAggCUUUUUUAAgCACACACCCAgUAUUgCCgUgACCCAUUACCgCAUgUgg
AUgACgCAUggAgACAAAAggAAAgUgAgCUgAAAAACCAgAgggAggAAAAgUggAAUggUgUgAUgCUgAgCgUUCAgAAAgUggCCAgAUgAgCUCAgAgUU
UCUgAUUUAAUgAAUgUAUgUgUgCgUgUgUgUgUggUUgggUCAUAUCUgAgACACUgUUCCACAgCAACAAAAACAAUAAAAUUCACUgUAUUUUCUCCUAAA
AAAAAAAAAAAAAAAAAAAAAA

Reaction pathway	Constrained sequence domains	Target function	Window length (nt)	Source mRNA	Candidate windows	Satisfying composition constraints	Satisfying pattern constraints	Satisfying composition and pattern constraints
Conditional self-assembly via hybridization chain reaction (HCR)	b*-a*	input X	36	*tpm3*	2140	565 (26%)	116 (5.4%)	89 (4.2%)
Boolean logic AND using toehold sequestration gates	a-b c-d	input X input Y	30 30	*tpm3* *desma*	2146 1769	641 (30%) 695 (39%)	190 (8.9%) 377 (21%)	114 (5.3%) 151 (8.5%)
Self-assembly of a 3-arm junction via catalytic hairpin assembly (CHA)	y*-b*-x*-a*	input X	36	*tpm3*	2140	565 (26%)	116 (5.4%)	89 (4.2%)
Boolean logic AND using a cooperative hybridization gate	a-b c-d	input X input Y	25 25	*tpm3* *desma*	2151 1774	510 (24%) 526 (30%)	298 (14%) 501 (28%)	115 (5.4%) 154 (8.7%)
Conditional Dicer substrate formation via shape and sequence transduction	a-b-c w-x-y-z	input X output Y	18 21	*tpm3* *desma*	2158 1778	317 (15%) 586 (33%)	576 (27%) 634 (36%)	117 (5.4%) 219 (12%)

Table S3. Window constraints for each reaction pathway based on mRNAs X and Y, as well as number of candidate windows satisfying: composition constraints ($\mathtt{S} \in [0.45, 0.55]$), pattern constraints (prevent $\{\mathtt{AAAA}, \mathtt{CCCC}, \mathtt{GGGG}, \mathtt{UUUU}, \mathtt{KKKKKK}, \mathtt{MMMMMM}, \mathtt{RRRRRR}, \mathtt{SSSSSS}, \mathtt{WWWWWW}, \mathtt{YYYYYY}\}$), or both.

S2.7 Robustness of Predictions to Model Perturbations

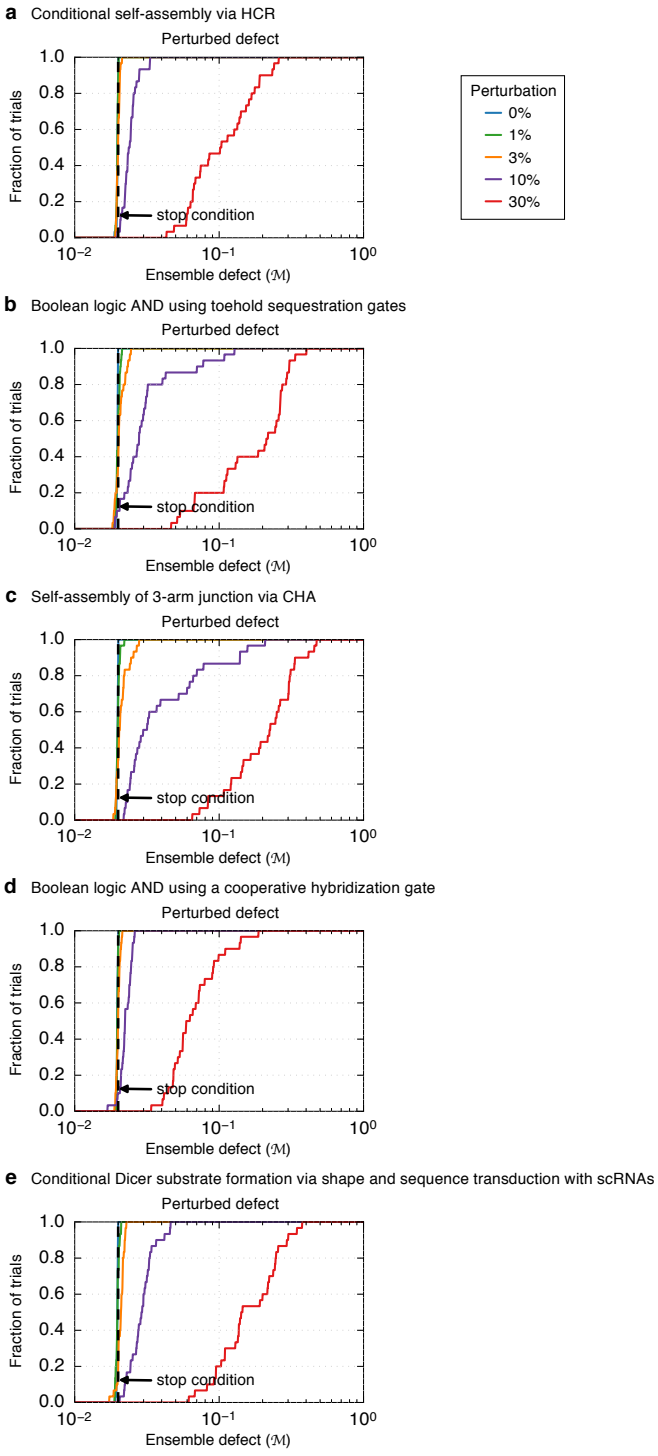


Figure S20. Robustness of design quality predictions to perturbations in model parameters ($N = 8$ orthogonal systems). For each design trial, the median multistate test tube ensemble defect was calculated over 100 perturbed physical models (each parameter perturbed by Gaussian noise with a standard deviation of 0, 1, 3, 10, or 30% of the parameter modulus). The stop condition is depicted as a dashed black line. (a) Conditional self-assembly via HCR. (b) Boolean logic AND using toehold sequestration gates. (c) Self-assembly of 3-arm junction via CHA. (d) Boolean logic AND using a cooperative hybridization gate. (e) Conditional Dicer substrate formation via shape and sequence transduction with scRNAs.

S3 Additional Design Studies

Here, we compare the performance of the current constrained multistate test tube design algorithm to that of the previously published test tube design algorithm[1] on the subsidiary design problems of test tube design and complex design (see Table 3 for a comparison of the design ensembles).

S3.1 Performance for Test Tube Design

Test tube design is a special case of multistate test tube design in which the design ensemble contains only one target test tube containing an arbitrary number of on-target and off-target complexes (Table 3). For test tube design comparisons, we use the *engineered test set* and *random test set* of target test tubes provided with Reference 1. For the engineered test set, each on-target structure was randomly generated with stem and loop sizes randomly selected from a distribution of sizes representative of the nucleic acid engineering literature. For the random test set, each on-target structure was generated by calculating the minimum free energy structure of a different random RNA sequence in 1 M Na^+ at 37 °C. Within each target test tube, there are two on-target dimers (each with a target concentration of 1 μM) and a total of 106 off-target monomers, dimers, trimers and tetramers (each with vanishing target concentration), representing all complexes of up to $L_{\max} = 4$ strands (excluding the two on-target dimers). For each test set, 50 target test tubes were generated for each on-target dimer size, $\{50, 100, 200, 400\}$ nt, with all strands the same length in each target test tube. The structural properties of the on-target structures in the engineered and random test sets are summarized in Supplementary Section S3 of Reference 1. Five design trials were run for each test case using stop condition $f_{\text{stop}} = 0.01$ (i.e., no more than 1% of nucleotides incorrectly paired at equilibrium). For test tube design, Figures S21 and S22 demonstrate that the performance of the current algorithm and the previously published test tube design algorithm[1] is similar on the engineered and random test sets.

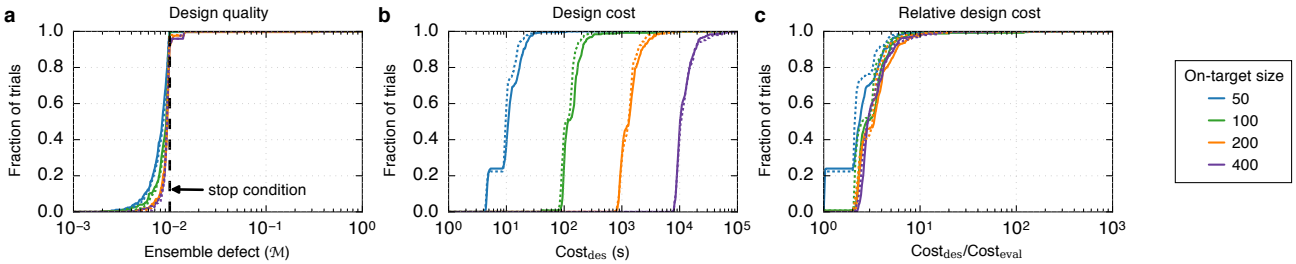


Figure S21. Algorithm performance for test tube design on the engineered test set. Comparison of the current multistate test tube design algorithm (solid lines) to the previously published test tube design algorithm[1] (dotted lines). (a) Design quality. The stop condition is depicted as a dashed black line. (b) Design cost. (c) Cost of sequence design relative to a single evaluation of the objective function. Design conditions: RNA in 1 M Na^+ at 37 °C.

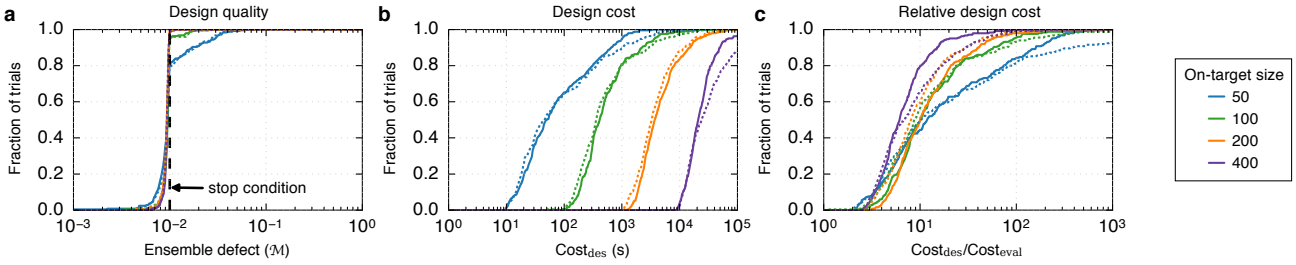


Figure S22. Algorithm performance for test tube design on the random test set. Comparison of the current multistate test tube design algorithm (solid lines) to the previously published test tube design algorithm[1] (dotted lines). (a) Design quality. The stop condition is depicted as a dashed black line. (b) Design cost. (c) Cost of sequence design relative to a single evaluation of the objective function. Design conditions: RNA in 1 M Na^+ at 37 °C.

S3.2 Performance for Complex Design

Complex design is a special case of multistate test tube design in which the design ensemble contains one target test tube containing one on-target complex and no off-target complexes (Table 3). For complex design comparisons, we use the (dimer) on-target structures from the target test tubes in the engineered and random test sets described above. Hence, for the studies, each target test tube contains a single on-target dimer and no off-targets; there are 50 on-target structures of each size $\{50, 100, 200, 400\}$ nt (using the first-listed on-target dimer in each of the 50 target test tubes of Section S3.1). Five design trials were run for each test case using stop condition $f_{\text{stop}} = 0.01$ (i.e., no more than 1% of nucleotides incorrectly paired at equilibrium). For complex design, Figures S23 and S24 demonstrate that the performance of the current algorithm and the previously published test tube design algorithm[1] is similar for the on-target structures in the engineered and random test sets.

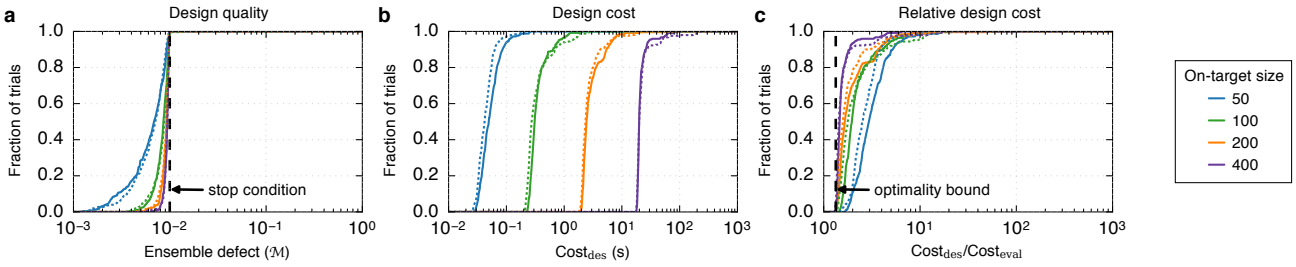


Figure S23. Algorithm performance for complex design using on-target structures from the engineered test set. Comparison of the current multistate test tube design algorithm (solid lines) to the previously published test tube design algorithm[1] (dotted lines). (a) Design quality. The stop condition is depicted as a dashed black line. (b) Design cost. (c) Cost of sequence design relative to a single evaluation of the objective function. The optimality bound[4] is depicted as a dashed black line. Design conditions: RNA in 1 M Na^+ at 37 °C.

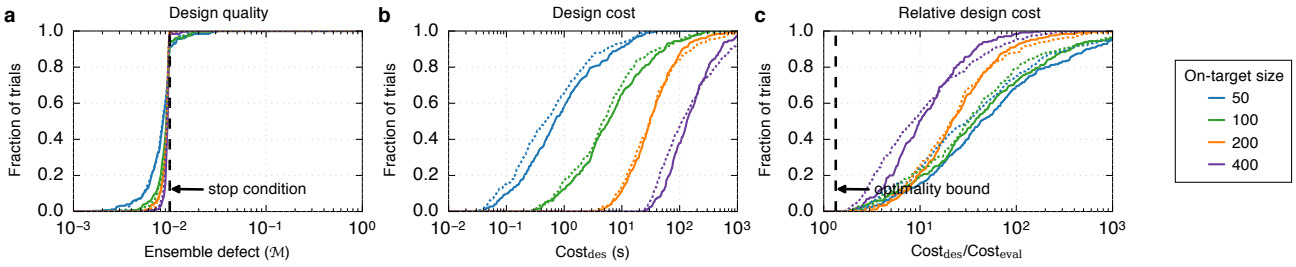


Figure S24. Algorithm performance for complex design using on-target structures from the random test set. Comparison of the current multistate test tube design algorithm (solid lines) to the previously published test tube design algorithm[1] (dotted lines). (a) Design quality. The stop condition is depicted as a dashed line. (b) Design cost. (c) Cost of sequence design relative to a single evaluation of the objective function. The optimality bound[4] is depicted as a dashed black line. Design conditions: RNA in 1 M Na^+ at 37 °C.

References

[1] Wolfe, B. R.; Pierce, N. A. *ACS Synth. Biol.* **2015**, 4, 1086–1100.

[2] Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. *SIAM Rev.* **2007**, 49, 65–88.

[3] Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. *J. Comput. Chem.* **2011**, 32, 170–173.

[4] Zadeh, J. N.; Wolfe, B. R.; Pierce, N. A. *J. Comput. Chem.* **2011**, 32, 439–452.

[5] Serra, M. J.; Turner, D. H. *Methods Enzymol.* **1995**, 259, 242–261.

[6] Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, 288, 911–940.

[7] Zuker, M. *Nucleic Acids Res.* **2003**, 31, 3406–3415.

[8] SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 1460–1465.

[9] SantaLucia, J., Jr.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, 33, 415–440.

[10] Koehler, R. T.; Peyret, N. *Bioinformatics* **2005**, 21, 3333–3339.

[11] Genot, A. J.; Zhang, D. Y.; Bath, J.; Turberfield, A. J. *J. Am. Chem. Soc.* **2011**, 133, 2177–2182.

[12] Genot, A. J.; Bath, J.; Turberfield, A. J. *J. Am. Chem. Soc.* **2011**, 133, 20080–20083.

[13] Delebecque, C. J.; Silver, P. A.; Lindner, A. B. *Nat. Protoc.* **2012**, 7, 1797–1807.

[14] Greene, D. G.; Keum, J. W.; Bermudez, H. *Small* **2012**, 8, 1320–1325.

[15] Padirac, A.; Fujii, T.; Rondelez, Y. *Nucleic Acids Res.* **2012**, 40.

[16] Tang, H.; Deschner, R.; Allen, P.; Cho, Y. J.; Sermas, P.; Maurer, A.; Ellington, A. D.; Willson, C. G. *J. Am. Chem. Soc.* **2012**, 134, 15245–15248.

[17] Zhang, X. J.; Yadavalli, V. K. *Nanoscale* **2012**, 4, 2439–2446.

[18] Goodman, D. B.; Church, G. M.; Kosuri, S. *Science* **2013**, 342, 475–479.

[19] Xu, X. W.; Yang, X. R. *Chem. Commun.* **2014**, 50, 805–807.

[20] Dirks, R. M.; Pierce, N. A. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 15275–15278.

[21] Patzel, V.; Rutz, S.; Dietrich, I.; Köberle, C.; Scheffold, A.; Kaufmann, S.H.E. *Nat. Biotechnol.* **2005**, 23, 1440–1444.

[22] Penchovsky, R.; Breaker, R.R. *Nat. Biotechnol.* **2005**, 23, 1424–1433.

[23] Venkataraman, S.; Dirks, R. M.; Rothemund, P. W. K.; Winfree, E.; Pierce, N. A. *Nat. Nanotechnol.* **2007**, 2, 490–494.

[24] Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, 451, 318–322.

[25] Salis, H. M.; Mirsky, E. A.; Voigt, C. A. *Nat. Biotechnol.* **2009**, 27, 946–950.

[26] Choi, H. M. T.; Chang, J. Y.; Trinh, L. A.; Padilla, J. E.; Fraser, S. E.; Pierce, N. A. *Nat. Biotechnol.* **2010**, 28, 1208–12.

[27] Li, B. L.; Ellington, A. D.; Chen, X. *Nucleic Acids Res.* **2011**, 39, e110.

[28] Choi, J.; Love, K. R.; Gong, Y.; Gierahn, T. M.; Love, J. C. *Anal. Chem.* **2011**, 83, 6890–6895.

[29] Dong, J.; Cui, X.; Deng, Y.; Tang, Z. *Biosens. Bioelectron.* **2012**, 38, 258–263.

[30] Nishimura, T.; Ogura, Y.; Tanida, J. *Appl. Phys. Lett.* **2012**, 101, 233703.

[31] Schade, M.; Knoll, A.; Vogel, A.; Seitz, O.; Liebscher, J.; Huster, D.; Herrmann, A.; Arbuzova, A. *J. Am. Chem. Soc.* **2012**, 134, 20490–20497.

[32] Vieregg, J. R.; Nelson, H. M.; Stoltz, B. M.; Pierce, N. A. *J. Am. Chem. Soc.* **2013**, 135, 9691–9699.

[33] Hochrein, L. M.; Schwarzkopf, M.; Shahgholi, M.; Yin, P.; Pierce, N. A. *J. Am. Chem. Soc.* **2013**, 135, 17322–17330.

[34] Genot, A. J.; Bath, J.; Turberfield, A. J. *Angew. Chem., Int. Ed.* **2013**, 52, 1189–1192.

[35] Hamblin, G. D.; Hariri, A. A.; Carneiro, K. M. M.; Lau, K. L.; Cosa, G.; Sleiman, H. F. *ACS Nano* **2013**, 7, 3022–3028.

[36] Santini, C. C.; Bath, J.; Tyrrell, A. M.; Turberfield, A. J. *Chem. Commun.* **2013**, 49, 237–239.

[37] Choi, H. M. T.; Beck, V. A.; Pierce, N. A. *ACS Nano* **2014**, 8, 4284–4294.

[38] Jiang, Y. S.; Bhadra, S.; Li, B. L.; Ellington, A. D. *Angew. Chem., Int. Ed.* **2014**, 53, 1845–1848.

[39] Geary, C.; Rothemund, P. W. K.; Andersen, E. S. *Science* **2014**, 345, 799–804.

[40] Green, A. A.; Silver, P. A.; Collins, J. J.; Yin, P. *Cell* **2014**, 159, 925–939.

[41] Hu, J. M.; Yu, Y. J.; Brooks, J. C.; Godwin, L. A.; Somasundaram, S.; Torabinejad, F.; Kim, J.; Shannon, C.; Easley, C. J. *J. Am. Chem. Soc.* **2014**, 136, 8467–8474.

[42] Machinek, R. R. F.; Ouldridge, T. E.; Haley, N. E. C.; Bath, J.; Turberfield, A. J. *Nat. Commun.* **2014**, 5, 5324.

[43] Franco, E.; Giordano, G.; Forsberg, P. O.; Murray, R. M. *ACS Synth. Biol.* **2014**, 3, 589–599.

[44] Koos, B.; Cane, G.; Grannas, K.; Löf, L.; Arngården, L.; Heldin, J.; Clausson, C. M.; Klaesson, A.; Hirvonen, M. K.; de Oliveira, F. M. S.; Talibov, V. O.; Pham, N. T.; Auer, M.; Danielson, U. H.; Haybaeck, J.; Kamali-Moghaddam, M.; Söderberg, O. *Nat. Commun.* **2015**, 6, 7294.

[45] Zalatan, J. G.; Lee, M. E.; Almeida, R.; Gilbert, L. A.; Whitehead, E. H.; La Russa, M.; Tsai, J. C.; Weissman, J. S.; Dueber, J. E.; Qi, L. S.; Lim, W. A. *Cell* **2015**, 160, 339–350.

[46] Galimidi, R. P.; Klein, J. S.; Politzer, M. S.; Bai, S. Y.; Seaman, M. S.; Nussenzweig, M. C.; West, A. P.; Bjorkman, P. J. *Cell* **2015**, 160, 433–446.

[47] Raschle, T.; Lin, C. X.; Jungmann, R.; Shih, W. M.; Wagner, G. *ACS Chem. Biol.* **2015**, 10, 2448–2454.

[48] Takahashi, M. K.; Watters, K. E.; Gasper, P. M.; Abbott, T. R.; Carlson, P. D.; Chen, A. A.; Lucks, J. B. *RNA* **2016**, 22, 920–933.

[49] Lee, Y. J.; Hoynes-O'Connor, A.; Leong, M. C.; Moon, T. S. *Nucleic Acids Res.* **2016**, 44, 2462–2473.

[50] Zuker, M.; Stiegler, P. *Nucleic Acids Res.* **1981**, 9, 133–147.

[51] McCaskill, J.S. *Biopolymers* **1990**, 29, 1105–1119.

[52] Hofacker, I.L.; Fontana, W.; Stadler, P.F.; Bonhoeffer, L.S.; Tacker, M.; Schuster, P. *Chem. Mon.* **1994**, 125, 167–188.

[53] Lyngsø, R. B.; Zuker, M.; Pedersen, C. N. S. *Bioinformatics* **1999**, 15, 440–445.

[54] Dirks, R. M.; Pierce, N. A. *J. Comput. Chem.* **2004**, 25, 1295–1304.

[55] Dimitrov, R.A.; Zuker, M. *Biophys. J.* **2004**, 87, 215–226.

[56] Andronescu, M.; Zhang, Z.C.; Condon, A. *J. Mol. Biol.* **2005**, 345, 987–1001.

[57] Bernhart, S.H.; Tafer, H.; Muckstein, U.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. *Algorithms Mol. Biol.* **2006**, 1.

[58] Dechter, R. **2003** *Constraint Processing* (Morgan Kaufmann, New York).

[59] Seelig, G.; Soloveichik, D.; Zhang, D.Y.; Winfree, E. *Science* **2006**, 314, 1585–1588.

[60] Zhang, D. Y. *J. Am. Chem. Soc.* **2011**, 133, 1077–1086.

[61] McEntyre, J.; Ostell, J. **2002** *The NCBI Handbook [Internet]* (National Center for Biotechnology Information (US), Bethesda, MD).